

Dirk Lewandowski & Christian Maaß (Hrsg.)

Web-2.0-Dienste als Ergänzung zu algorithmischen Suchmaschinen

II



Das vorliegende Buchprojekt entstand im Rahmen des Forschungsprojektes Theseus (Teilprojekt Alexandria), das unter dem Förderkennzeichen 01MQ07008 vom BMWi gefördert wird.

Theseus im Internet: <http://www.theseus-programm.de>

Inhaltsverzeichnis

Vorwort	VII
----------------------	------------

Suchmaschinen: Status quo und Entwicklungstendenzen.....	1
---	----------

1. Einleitung	1
2. Entwicklung und wirtschaftliche Bedeutung algorithmischer Suchmaschinen.....	3
3. Charakterisierung und Marktüberblick alternativer Suchdienste.....	5
3.1 Soziale Bookmarkdienste	6
3.2 Frage-Antwort-Dienste.....	7
3.3 Semantische Suchdienste	8
3.4 Konvergente Suchdienste	9
4. Fazit	10
Verwendete Literatur	11

Qualitative Laboruntersuchung zur Anwendung von Social-Software-Systemen.....	13
--	-----------

1. Einleitung	13
2. Qualitative Laboruntersuchung.....	14
2.1 Technischer Ablauf	14
2.2 Durchführung	15
2.3 Auswertung der Laboruntersuchung	18
2.4 Suche allgemein	18
2.5 Situationsaufgabe 1: Meinungs- und Empfehlungsfragen.....	21
2.6 Situationsaufgabe 2: Faktenfragen	22
2.7 Situationsaufgabe 3: Bookmarks.....	22
2.8 Einstellen einer eigenen Frage	26
2.9 Besonderheiten	30
2.10 Weitere Nutzung von Lycos iQ	32
3. Fazit	35
Verwendete Literatur	37
Anhang	38
Anhang 1 – Fragebogen: Hintergrundwissen der Testperson.....	38
Anhang 2 – Auswertung des Fragebogens	40
Anhang 2a – Profil der Testpersonen	42

IV

Anhang 3 – Ablauf für die aktive Laboruntersuchung	43
Anhang 4 – Einverständniserklärung	47
Anhang 5 – Testaufgaben.....	48
Anhang 6 – Feedbackbogen Lycos iQ	50
Anhang 7 – Auswertung der Feedbackbögen.....	51

Vergleich der Relevanz von Treffern bei algorithmischen Suchmaschinen, Social Bookmarking-Seiten und Frage-Antwort-Diensten 55

1. Einleitung	55
2. Methodik	56
2.1 Vorüberlegung.....	56
2.2 Voruntersuchung (Pre-Test).....	57
2.2.1 Selektion der Fragen	58
2.2.2 Vorgehensweise	58
2.2.3 Auswertung	58
2.2.4 Ergebnisse.....	59
2.3 Haupttest (Main-Test)	60
2.3.1 Suchanfragen	60
2.3.2 Vorgehensweise und Bewertung der Probanden.....	61
3. Auswertung des Haupttests.....	61
3.1 Top Ten Precision	62
3.2 Weitere Bewertungskriterien.....	67
3.2.1 Verständlichkeit	67
3.2.2 Vertrauenswürdigkeit.....	69
3.2.3 Wikipedia-Seiten	69
3.2.4 Ausreichende Information	70
3.3 Überschneidungen.....	71
3.4 Irrelevante Treffer	75
4. Fazit	79
Verwendete Literatur.....	80
Weiterführende Literatur	81

Untersuchung der Qualität der Antworten bei Lycos iQ und deren Einbindung in die algorithmische Suche 83

1. Einleitung	83
2. Methodische Vorgehensweise	84
2.1 Einteilung der Antworten in Themengruppen.....	85

2.2 Erstellung des Kriterienkatalogs	85
3. Auswertung der Fragenkataloge	90
3.1 Zuweisung der Themengebiete	90
3.2 Probleme nach der Themenzuweisung.....	90
3.3 Auswertung des Faktenfragenkatalogs.....	92
3.4 Auswertung des Erfahrungsfragenkatalogs.....	109
4. Fazit	121
4.1 Die Qualität der Antworten	121
4.2 Zusammenhang zwischen Expertenstatus und Qualität der Antworten ..	125
4.3 Kriterien zur Einbindung der Antworten in die algorithmische Suche...	126
4.4 Allgemeine Anmerkungen	127
Verwendete Literatur.....	128
 Analyse der Fragensituation bei Lycos iQ	 129
1. Einleitung	129
2. Thesen	130
3. Methodik und Vorüberlegungen	130
3.1 Thema der Fragen.....	130
3.2 Art der Fragen	131
4. Arbeitsschritte.....	135
5. Kriterienkatalog zur Bewertung der Antworten	135
5.1 Die Untersuchung der Antworten nach festgelegten Qualitätsmerkmalen	136
5.2 Die Bewertung der Antworten nach den Qualitätskriterien anhand des Punktesystems	138
6. Methodik: Fragekatalog	143
6.1 Formale Kriterien	143
6.2 Inhaltliche Kriterien	144
6.3 Schwierigkeitsgrad	145
6.4 Sonstige Kriterien.....	146
7. Methodik der Auswertung.....	147
8. Ergebnisse	148
8.1 Verteilung der Antwortqualität	148
8.2 Schwierigkeitsgrad	150
8.3 Anzahl der Antworten	153
8.4 Bonuspunkte.....	154
8.5 Gut-Bewertungen	155
8.6 Formale Merkmale	157
8.7 Inhaltliche Merkmale	158

9. Einbinden von Fragen aus Lycos iQ in andere Suchformen.....	159
10. Tutorial.....	161
11. Fazit	161

Verbesserung des Taggings in Lycos iQ..... 163

1. Einleitung	163
2. Probleme beim Tagging	163
2.1 Synonyme.....	164
2.2 Homonyme.....	164
2.3 Abkürzungen / Akronyme.....	164
2.4 Singular- und Pluralform.....	164
2.5 Rechtschreibung.....	164
2.6 Zusammengesetzte Begriffe.....	165
2.7 Spezifizierung	165
2.8 Sprache.....	165
3. Kriterien für die Analyse der unterschiedlichen Dienste	166
4. Analyisierte Dienste	167
4.1 Frage-Antwort-Dienste.....	168
4.2 Social Bookmarking-Dienste	172
4.3 Andere Web-Anwendungen	177
5. Vorschläge für Lycos iQ.....	184
5.1 Erweiterung der Tagging-Rechte	185
5.2 Verbesserte Editiermöglichkeit der Tags	185
5.3 Cluster-Funktion	186
5.4 Motivation.....	186
5.5 Verwaltung der Tags	186
5.6 „Suggestive tagging“-Funktion	186
5.7 Flash Tutorials.....	187
5.8 Rechtschreibhilfe.....	187
5.9 Alternative zur Rechtschreibhilfe.....	187
5.10 Kategorien	188
6. Fazit	189
Verwendete Literatur.....	189

Vorwort

DIRK LEWANDOWSKI

dirk.lewandowski@haw-hamburg.de

CHRISTIAN MAASS

christian@ebusiness-management.info

Der vorliegende Band dokumentiert die Ergebnisse des Fachprojektes „Einbindung von Frage-Antwort-Diensten in die Web-Suche“, welches im Wintersemester 2007/2008 am Department Information der Hochschule für Angewandte Wissenschaften Hamburg in Zusammenarbeit mit Lycos Europe und der Online-Sparte der Deutschen Telekom, T-Online, durchgeführt wurde.

Mit den Frage-Antwort-Diensten ist eine Ergänzung, teilweise sogar eine Konkurrenz zu den bisherigen Ansätzen in der Web-Suche entstanden. Während Google und andere Suchmaschinen automatisch generierte Trefferlisten bieten, die dem Nutzer relevante Dokumente anzeigen, gehen die Frage-Antwort-Dienste darüber hinaus: Sie binden die Nutzer sowohl als Fragesteller als auch als Antwortgeber ein und liefern gleichzeitig direkte Antworten auf die Fragstellungen geben, ohne dem Nutzer das Finden der gewünschten Antwort innerhalb von Dokumenten aufzubürden.

Als Beispiele für Frage-Antwort-Dienste sind im deutschsprachigen Bereich vor allem Yahoo! Clever und Lycos iQ zu nennen. Beiden Diensten ist es gelungen, große Communities aufzubauen, in denen sich Nutzer anhand der gestellten Fragen austauschen. Es liegt in der Natur der Sache, dass diese Dienste Antworten erst mit einer gewissen Verzögerung nach dem Stellen einer Frage geben können – verwundert hat uns jedoch, wie schnell oft die ersten Antworten eintreffen. Dieses Beispiel zeigt, dass es sich bei Community-basierten Diensten nicht allein um technische Systeme handelt, sondern vielmehr auch der Aufbau der Gemeinschaft und ihre Pflege eine wichtige Rolle spielen. Insofern sind die in diesem Band enthaltenen Kapitel keine Berichte über die technische Weiterentwicklung der Frage-Antwort-Dienste, sondern vielmehr eine kritische Beleuchtung einzelner Aspekte dieses Phänomens, die ebenfalls eine umfassende Qualitätsbewertung der Dienste liefern.

Das Kapitel von Andre Skusa und Christian Maaß leitet das Buch mit einem Überblick über die verschiedenen Arten von Suchdiensten im Internet ein und diskutiert deren Stärken und Schwächen. Dabei wird auf algorithmische, soziale, semantische und verschiedene Mischformen dieser Suchdienste eingegangen. Im Zuge dieser Diskussion wird die Frage aufgestellt, inwieweit die Ergebnisqualität von Suchmaschinen durch eine Kombination verschiedener Suchdienste gesteigert werden kann und welche Bedeutung in diesem Zusammenhang soziale Suchdienste spielen, die man gerade in jüngerer Zeit immer häufiger als alternative Suchansätze in der Presse diskutiert. Diese Frage wird in den Folgebeiträgen aus unterschiedlichen Facetten weiter beleuchtet.

VIII

In der qualitativen Laboruntersuchung von Anneliese Bergen und Tina Krause wird anhand eines ausführlichen Experiments gezeigt, wie Nutzer mit den verschiedenen Suchdiensten umgehen und sich ihrer Meinung nach die unterschiedlichen Suchansätze ergänzen lassen. Die Ergebnisse zeigen, dass den Nutzern die neuen Frage-Antwort- sowie Social Bookmarking-Angebote (noch) nicht bekannt sind und sie auch durch ihre bisherige Ergebnisqualität noch nicht vollständig überzeugen können.

Der Ergebnisqualität von algorithmischen Suchmaschinen, Social-Bookmarking-Systemen und Frage-Antwort-Diensten widmet sich der Aufsatz von Olga Gammer, Heidi Meißner, Magdalena Preckel und Robert Oehlert unter der klassischen informationswissenschaftlichen Fragestellung nach der Trefferrelevanz. Sie führen einen Test zur Retrievaleffektivität durch und vergleichen die ausgegebenen Ergebnisse im Rahmen einer anonymisiert durchgeführten Untersuchung. Dabei zeigt sich, dass gegenwärtig die algorithmischen Suchmaschinen im Durchschnitt die besten Treffer liefern, während Social Bookmarking-Dienste ihr Potential noch nicht erfüllen können. Sie liefern trotz der vorherigen Empfehlung durch die Nutzer auch schlechtere Ergebnisse als die Frage-Antwort-Dienste, die den mittleren Rang belegen. Weiterhin auffällig ist, dass sich die Suchdienste-Typen hinsichtlich der Relevanz klar voneinander abgrenzen, also kein einzelnes Suchangebot so gut oder schlecht ist, dass es in eine der anderen Gruppen auf- oder absteigt.

Der Beitrag von Konstantin Guratzsch, Sonja von Mach, Jessica Otte und Joël Pientok beschäftigt sich mit der Integration von Antworten aus Frage-Antwort-Diensten in die Ergebnislisten von algorithmischen Suchmaschinen. Auch hier stellt sich wieder die Frage nach der Qualität: Es sollten natürlich nur solche Antworten eingebunden werden, die sowohl faktisch richtig sind als auch gewissen formalen Kriterien genügen. Als Untersuchungsgegenstand wurde hier Lycos iQ ausgewählt. Es zeigt sich, dass der Anteil der richtigen Antworten bei über 90 Prozent liegt und dass sich das verwendete Reputationssystem für das Einblenden passender Antworten verwenden lässt.

Von der anderen Seite, nämlich von der Qualität der Fragen her, nähert sich die Untersuchung von Philipp Niehues, Diana Schulz, Vanessa Schäfer und Bianca Omelan dem Thema. Sie beschreiben einen Kriterienkatalog für die Bewertung der Fragen hinsichtlich ihrer Eignung für die Einbindung in die Suche. Da gut gestellte Fragen potentiell auch gute Antworten generieren, ist es wichtig, Kriterien zu finden, nach denen Fragen eingebunden werden sollen. Von Bedeutung ist hier die Unterscheidung zwischen Fakten- und Erfahrungsfragen, die in der Einbindung verschieden behandelt werden sollten.

Das abschließende Kapitel von Stefan Seehaus und Fabian Lemm beschäftigt sich mit dem Tagging in Lycos iQ, also der freien Verschlagwortung der Fragen und Antworten, die eine bessere Auffindbarkeit der archivierten Inhalte garantieren soll. Dazu wird ein umfassender Überblick über andere, teils grundverschiedene Dienste im Web gegeben und Best Practices für Lycos iQ (bzw. andere Angebote, die ein möglichst optimales Tagging verwenden wollen) abgeleitet.

In seiner Gesamtheit beleuchtet der vorliegende Band die Frage-Antwort-Dienste aus unterschiedlichen Blickwinkeln. In den einzelnen Beiträgen können zahlreiche Ansatzpunkte für weitere wissenschaftlichen Forschungen, aber auch für die Praxis gegeben werden. Auch über den direkten Anwendungsfall Frage-Antwort-Dienste

hinaus sind die Ergebnisse verwendbar für Entwickler von Community-basierten Diensten. Wir sind davon überzeugt, mit den vorgelegten Ergebnissen einen Startpunkt für weitere Forschungen und Entwicklungen geben zu können. Für Anregungen und Kritik sind wir allen Leserinnen und Lesern dankbar; bitte kontaktieren Sie uns direkt über unsere am Beginn dieses Vorworts angegebenen E-Mail-Adressen.

Nicht nur die zu den einzelnen Beiträgen genannten Autorinnen und Autoren haben zum Erfolg des Projekts und der vorliegenden Publikation beigetragen. Wir möchten uns herzlich bei Meike Paustian, Jan Hoenck, Jan van Hamme, Rembert Wohlers (alle HAW Hamburg), Jens Fauldrath, Benjamin Walter (beide T-Online), Elica Savova, Isabelle Adelt, Karl Szwillus, Thomas Dominikowski (alle Lycos Europe) für ihre Mitarbeit in unterschiedlichen Rollen bedanken.

Hamburg, Gütersloh, im Juni 2008

Prof. Dr. Dirk Lewandowski, Hochschule für Angewandte Wissenschaften
Hamburg

Dr. Christian Maaß, Lycos Research

Suchmaschinen: Status quo und Entwicklungstendenzen

ANDRE SKUSA

andre.skusa@lycos-europe.com

CHRISTIAN MAASS

christian@ebusiness-management.info

Suchmaschinen erfassen und sortieren die Inhalte des Internets und sind über die letzten Jahre hinweg zu den meistgenutzten Webanwendungen avanciert. Im Zuge der Auseinandersetzung mit dem Suchmaschinenmarkt stehen dabei häufig Unternehmen wie Google, Yahoo! oder Microsoft im Vordergrund. In jüngerer Zeit lässt sich jedoch beobachten, dass neben diesen algorithmenbasierten Suchmaschinen immer häufiger alternative Suchdienste in den Markt treten, wie z. B. soziale, semantische oder so genannte konvergente Suchdienste. In diesem Beitrag wird vor diesem Hintergrund aufgezeigt, dass diese Suchdienste zum jetzigen Zeitpunkt noch nicht mit Google und anderen führenden Suchmaschinen konkurrieren können. Es zeigt sich jedoch, dass eine Kombination verschiedener Suchdienste zu einer Verbesserung der Ergebnisqualität von Suchmaschinen beitragen kann.

1. EINLEITUNG¹

Mit dem exponentiellen Wachstum des Internets sind Suchmaschinen wie Google und Yahoo! zu den meistgenutzten Internetanwendungen avanciert, die von etwa 90 Prozent der Internetnutzer in Deutschland regelmäßig zu Recherchezwecken genutzt werden. Insbesondere im Vorfeld von Kaufentscheidungen spielen sie – sowohl für Unternehmen als auch für Privathaushalte – eine immer wichtigere Rolle, um gezielt nach bestimmten Produkten und Dienstleistungen zu recherchieren. Zur Befriedigung ihrer Informationsbedürfnisse können die Anwender grundsätzlich auf eine Vielzahl verschiedener Suchmaschinen zurückgreifen; allein für den deutschsprachigen Raum sind weit über hundert verschiedene Suchmaschinen verfügbar (vgl. für einen Überblick z. B. www.mysuche.de). Mit Blick auf die Praxis muss jedoch ein Oligopol algorithmenbasierter Suchmaschinen – bestehend aus Google, Yahoo! und MSN – konstatiert werden, das den weltweiten Markt dominiert; unter algorithmischer Suche versteht man dabei ein automatisiertes Verfahren, das im Internet vorhandene Seiten erfasst, mit Hilfe statistischer Methoden bewertet, sortiert und zu den Worten der Suchanfrage in Beziehung setzt.

Die Vormachstellung von diesen Suchmaschinen wird in der Öffentlichkeit bereits seit geraumer Zeit kritisiert. Teilweise wird in diesem Zusammenhang darauf hingewiesen, dass aufgrund dieser Marktkonstellation eine Gefahr für die Informationsvielfalt und -auswahl im Internet besteht, die nahezu durch eine Handvoll

¹ Dieser Artikel entstand unter Mithilfe von Jan Hoenck, Jan van Hamme und Rembert Wohler.

2 Status quo und Entwicklungstendenzen

Unternehmen mit kommerziellen Interessen beeinflusst werden [1]. Zu ähnlichen Äußerungen kommt es seitens der Politik [2, 3], die bereits seit längerer Zeit auf die Monopolisierungstendenzen im Suchmaschinenmarkt hinweist und sich für ein größeres Angebot alternativer Suchdienste ausspricht. In jüngerer Zeit kann man tatsächlich beobachten, dass in zunehmendem Maße Suchmaschinen mit alternativen Suchansätzen in den Markt treten. Möglichkeiten alternativer Suchansätze bestehen beispielsweise darin, das Wissen menschlicher Nutzer in den Suchprozess einzubeziehen (soziale Suchdienste) oder die bisherigen algorithmenbasierten Verfahren durch semantische Technologien zu verbessern (semantische Suchdienste). Weiterhin sind auch Kombinationen der hier genannten Suchansätze denkbar (konvergente Suchdienste), wie z. B. die Bewertung und Korrektur automatisch erzeugter Suchergebnisse durch Benutzer. Exemplarisch hierfür sei die Suchmaschine Wikia genannt; Abbildung 1.1 zeigt die hier angesprochenen Suchdienste im Überblick.

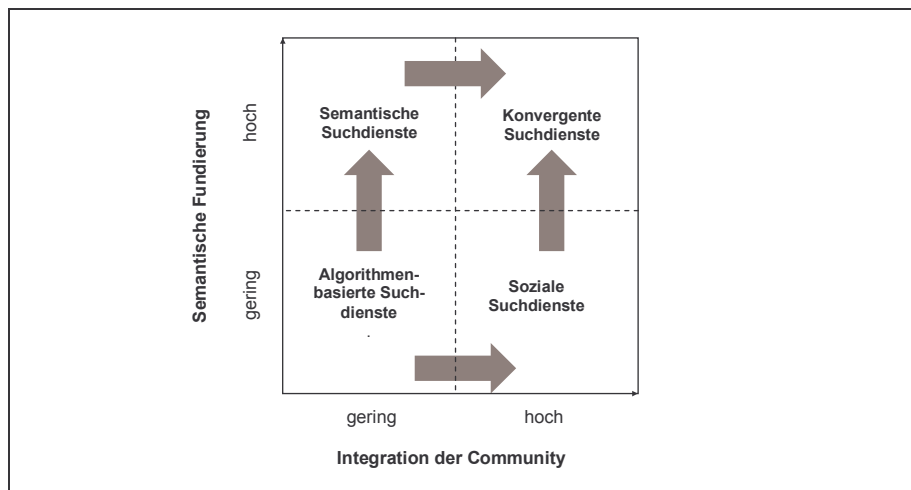


Abb. 1.1 Einteilung unterschiedlicher Typen von Suchdiensten

In Anbetracht der hier skizzierten Entwicklungstendenzen drängt sich unweigerlich die Frage auf, welche Bedeutung alternative Suchdienste im Allgemeinen und die aus der Kombination resultierenden konvergenten Suchdienste im Speziellen haben. Im weiteren Verlauf dieses Beitrags wird diese Frage ausführlicher beleuchtet. Um für die Auseinandersetzung mit diesem Thema zu sensibilisieren, gilt es zunächst jedoch, die Entwicklung und wirtschaftliche Bedeutung des Suchmaschinenmarktes aufzuarbeiten und die in diesem Kontext diskutierten algorithmischen Suchmaschinen zu skizzieren. Auf dieser Grundlage erfolgt anschließend eine Auseinandersetzung mit alternativen Suchdiensten. Dabei wird aufgezeigt, dass in der Verknüpfung von algorithmenbasierten und sozialen Suchdiensten ein interessanter, aber weitestgehend unerforschter Ansatz zu sehen ist, um die Qualität der Suchmaschinenergebnisse zu verbessern. Damit wird gleichzeitig die konzeptionelle Grundlage für den weiteren Verlauf dieses Buches gelegt.

2. ENTWICKLUNG UND WIRTSCHAFTLICHE BEDEUTUNG ALGORITHMISCHER SUCHMASCHINEN

Die Entwicklung des Suchmaschinenmarktes ist vergleichsweise jung und eng mit der Entwicklung des Internets und dessen wichtigstem Dienst, dem *World Wide Web* (WWW) verbunden. Der Grundstein des WWW wurde im Jahr 1989 am schweizerischen Kernforschungsinstitut CERN gelegt. Hier arbeitete man daran, Dokumente elektronisch miteinander zu verknüpfen, um auf diesem Wege den Informationsaustausch zwischen den Wissenschaftlern zu forcieren. Die Zahl dieser so genannten Hypertexte war bereits nach kurzer Zeit für Menschen nicht mehr überschaubar und es entstand ein Bedarf nach Hilfsmitteln, um die Hypertexte zu durchsuchen. Nachdem mit Archie bereits im Jahr 1990 ein Suchroboter für FTP-Dateien zur Verfügung stand, entstand im Frühjahr 1993 der erste Suchroboter für das WWW, der World Wide Web Wanderer. Er durchsuchte zwischen 1993 und 1996 zwei Mal pro Jahr das weltweite Datennetz [4]. Mit der Verbreitung solcher Suchmaschinen wurden die Verzeichnisdienste zunehmend abgelöst, in denen mittels manueller Zuordnung das WWW katalogisiert wurde; die heute noch bekannte Suchmaschine Yahoo! startete 1995 als ein solcher Verzeichnisdienst.

Ab Mitte der 90er Jahren entstanden mit dem zunehmenden Wachstum des Internets vor allem im universitären Umfeld weitere Suchmaschinen. Der im Jahr 1994 entwickelte *RBSE Spider* und der an der Universität von Washington entstandene Webcrawler waren z. B. die ersten algorithmischen Suchtechnologien, die ihre Ergebnisse bewertet haben (Ranking) und sie nach dieser Bewertung sortierten. Etwa zeitgleich entstand an der Carnegie Mellon University die Suchtechnologie Lycos, die in ihrer Relevanzbewertung auch die „word proximity“ berücksichtigte – also die Nähe verschiedener Suchwörter zueinander – und damit schnell Marktanteile gewinnen konnte.

Den bislang skizzierten algorithmischen Suchmaschinen ist gemein, dass sie die im Web auffindbaren Daten mit Hilfe von Crawlern sammeln, automatisch analysieren und dem Anwender den Zugriff auf diesen Index in Form einer Webseite ermöglichen. Nach diesem Prinzip funktioniert grundsätzlich auch die 1998 gegründete Suchmaschine Google, die diese Aufgaben jedoch wesentlich besser als ihre Konkurrenten bewältigte. Genauer gesagt gelang Google ein Durchbruch bei der Indizierung der im Web auffindbaren Dokumente. Während die seinerzeit etablierten Suchmaschinen wie Lycos und AltaVista die Relevanz ihrer Ergebnisse hauptsächlich anhand der im Dokument auffindbaren Schlüsselwörter bestimmten, integrierte Google zusätzlich den so genannten PageRank, nach dem die Relevanz eines Dokumentes rekursiv von den Verweisen auf dieses Dokument abhängt. Das bedeutet, dass eine Seite umso höher bewertet wird, je mehr andere Seiten auf sie verweisen, d.h. je bekannter sie ist. Diesem Ansatz liegt die Annahme zugrunde, dass auf Webseiten mit qualitativ guten Inhalten wesentlich mehr Links verweisen als auf inferiore Webangebote. Nicht zuletzt dieses Verfahren verhalf Google – trotz des relativ späten Markteintritts 1998 – zum Durchbruch und dem bis heute stetig wachsenden Marktanteils. Er beträgt in den USA etwa 61 Prozent, während Yahoo! und MSN mit 20,4 Prozent und 9,1 Prozent deutlich abgeschlagen sind (vgl. Abb. 1.2). Einige einst bekannte Suchmaschinen wie Alta Vista haben nahezu keine Marktbedeutung mehr. Noch drastischer fällt die Vormachstellung von Google

4 Status quo und Entwicklungstendenzen

allerdings in Deutschland aus, wo das Unternehmen einen Marktanteil von rund 90 Prozent aufweisen kann.

	Marktanteil im April 2008	Anzahl der Suchanfragen im April 2008 (in Millionen)
Google	61,6 %	6.514
Yahoo!	20,4 %	2.159
Microsoft	9,1 %	961
AOL	4,6 %	491
Ask.com	4,3 %	458

Abb 1.2 Marktanteile im deutschen Suchmaschinenmarkt [5]

2.1 Wirtschaftliche Bedeutung

Sind Werbeeinnahmen in der Internetbranche generell eine der Haupteinnahmequellen, so trifft dies auf Suchmaschinen in noch stärkerem Maße zu. So zählt die Suchmaschinenwerbung gegenwärtig zu den beliebtesten Werbeformen im Internet und weist im Vergleich zu traditionellen Werbeformen überdurchschnittliche Wachstumsraten auf. Allein in Deutschland wurden im Jahr 2007 mit dieser Werbeform knapp 1,2 Milliarden Euro und damit fast 40 Prozent mehr als im Vorjahr umgesetzt (vgl. auch Abb. 1.3). Mittelfristig rechnet man sogar damit, dass die Suchmaschinenwerbung – ähnlich wie in den USA – deutlich mehr als 50 Prozent der gesamten Online-Werbeausgaben auf sich vereinigt; die Schätzungen bezüglich des Volumens des weltweiten Online-Werbemarktes liegen zwischen 20 und 50 Milliarden Euro [6, 7]. Insofern ist auch in Zukunft damit zu rechnen, dass algorithmische Suchmaschinen weiterhin hohe Gewinne erwirtschaften.

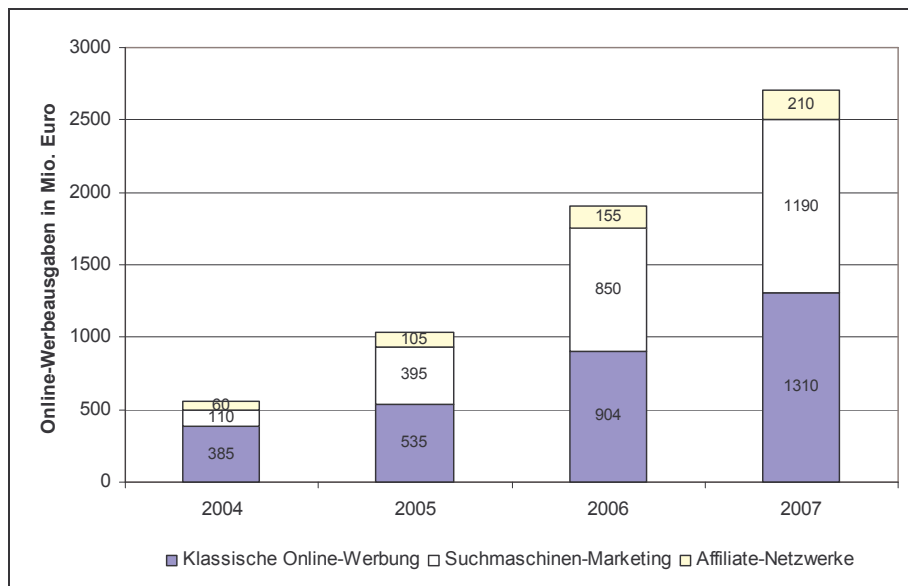


Abb. 1.3 Entwicklung des deutschen Online-Werbemarktes [8]

Die Funktionsweise des eigentlichen Erlösmodells von Suchmaschinen basiert vor diesem Hintergrund auf einem relativ einfachen Prinzip: Die Werbetreibenden müssen zur Platzierung ihrer Werbeanzeigen lediglich bestimmte Schlüssel- bzw. Suchwörter festlegen, unter denen die Anzeigen erscheinen sollen. Die Anzeigen als solche verweisen dann auf deren Internetpräsenzen, wobei für jeden Klick auf die Anzeige ein bestimmter Preis an den Suchmaschinenbetreiber zu entrichten ist; diese Form der Abrechnung wird im Allgemeinen als „Pay per Click“ bezeichnet. Bei Suchmaschinenwerbung werden dem Anwender folglich nur die zu seiner Suchanfrage passenden Anzeigen präsentiert. Dadurch steigt im Vergleich zu unspezifischen Werbeeinblendungen die Wahrscheinlichkeit, dass der Anwender auf die Anzeige reagiert.

Auf Basis des hier skizzierten Erlösmodells erwirtschaften die etablierten Suchmaschinen Gewinne in Milliardenhöhen. Allen voran ist hier das Unternehmen Google zu nennen, dessen Erlösmodell zu 99 Prozent auf so einer Werbefinanzierung basiert, womit jedes Quartal ein Gewinn von etwa 1 Milliarde US-Dollar einhergeht [9]. Die hohe Akzeptanz von Google liegt vor allem darin begründet, dass neben der neuartigen Bewertung der Suchergebnisse (PageRank, s. o.) und dem puristischen, auf das notwendigste beschränkten Design, die eingeblendete Werbung zu den Suchergebnissen passend und unauffällig ist. So kann diese Zusatzinformation zwar genutzt werden, ist aber trotzdem eindeutig als Werbung erkennbar. So wird der Nutzer nicht vom eigentlichen Suchergebnis abgelenkt, der Betreiber aber kann aufgrund der Beliebtheit der Seite genug aus den Werbeeinnahmen Erlösen.

Aufgrund der hier skizzierten wirtschaftlichen Potenziale erstaunt es kaum, dass alternative Suchdienste immer stärker in den Markt drängen, um vom lukrativen Geschäft mit der Suchmaschinenwerbung zu profitieren. Sowohl in der Wirtschaftspresse als auch in der Wissenschaft rückt dabei die Auseinandersetzung mit

- sozialen Suchdiensten
- semantischen Suchdiensten und
- konvergenten Suchdiensten

immer mehr in den Vordergrund. Diese Ansätze werden im weiteren Verlauf ausführlicher beleuchtet.

3. CHARAKTERISIERUNG UND MARKTÜBERBLICK ALTERNATIVER SUCHDIENSTE

Neben algorithmischen Suchmaschinen erfreuen sich in jüngerer Zeit auch so genannte soziale Suchdienste zunehmender Beliebtheit. Dabei handelt es sich um spezifische Ausprägungsformen sozialer Softwarelösungen, die in einer allgemeinen Begriffsumschreibung auf den Aufbau von sozialen Netzwerken sowie die Publikation und Verteilung von Informationen innerhalb eines Netzwerkes abstellen. „Social-Software-Systeme sind [...] umfassende sozio-technische Systeme, die auf Basis technischer und sozialer Vernetzung durch einfach zu bedienende Informationssysteme gemeinsam in einem bestimmten Themenfeld Leistungen

generieren“ [10]. Bei sozialen Suchdiensten bestehen diese Leistungen darin, dass die Informationsbedürfnisse der Anwender nicht von Suchmaschinenrobotern, sondern von einer Community aus Freiwilligen erbracht werden, die sich in erster Linie aus altruistischen Motiven bei diesen Diensten einbringen. Exemplarisch für solche Suchdienste wird im Folgenden auf

- soziale Bookmarksysteme und
- Frage-Antwort-Dienste eingegangen.

3.1 Soziale Bookmarkdienste

Soziale Bookmarkdienste wie z. B. del.icio.us, Furl oder Mister Wong – werden teilweise bereits als Alternative für die algorithmischen Suchmaschinen. gehandelt [11]. Diese Dienste bieten den Benutzern die Möglichkeit ihre persönlichen Lesezeichen (Bookmarks) im Netz zu speichern, zu verwalten und anderen Nutzern zur Verfügung zu stellen. Einmal gespeichert, können diese gern besuchten, nützlichen und persönlich ausgesuchten Webseiten jederzeit wieder gefunden und von vielen genutzt werden. Dies ermöglicht eine Suche in einer von menschlichen Anwendern zusammengestellten Treffermenge anstatt in Ergebnissen eines statistischen Bewertungsverfahrens. Die Funktionsweise dieser Dienste beruht dabei im Wesentlichen auf Tags [12]. Hierbei handelt es sich um frei wählbare Stichworte, die den Inhalt der gespeicherten Seite möglichst treffend charakterisieren sollen. Die zugeordneten Tags werden genutzt, um die auf den Seiten enthaltene Information zu beschreiben und auffindbar zu machen. Durch diese Tags entsteht im Laufe der Zeit eine Folksonomy [12], also eine durch eine große Menge von Personen definierte Klassifikation von Internetadressen. Anders formuliert können auf diesem Wege inhaltlich wertvolle, aber statistisch irrelevante Informationen gefunden werden, sofern sie von anderen Benutzern in den gemeinschaftlich gepflegten Index aufgenommen und sinnvoll annotiert werden. Dies führt unmittelbar zu der Frage, inwieweit eine Kombination dieser Suchdienste mit algorithmenbasierten Suchmaschinen zu einer Verbesserung der Ergebnisqualität beitragen könnte (vgl. hierzu insbesondere den Beitrag von Gammer u. a. in diesem Band).

Die gegenwärtig am Markt existierenden Bookmarksysteme – wie z. B. Mister Wong, del.icio.us, Furl, OneView, Bibsonomy, CiteULik, LinkArena usw.. – unterscheiden sich vor diesem Hintergrund nicht gravierend voneinander und haben ähnliche Features. Ebenso arbeiten alle Dienste mit freiem Tagging und bieten die Möglichkeit, bereits bestehende Lesezeichensammlungen zu importieren und die beim Dienst gespeicherten Lesezeichen zu exportieren. RSS-Feeds und die Integration in andere Tools können bei jedem Anbieter in Anspruch genommen werden. Einige Anbieter stellen ihren Anwendern jedoch besondere Dienste zur Verfügung: So werden bei Furl z. B. Kopien ganzer Webseiten gespeichert. Mit Oneview lassen sich hingegen auch Artikel und Videos verwalten. Ähnlich funktioniert auch Linkarena, wo auf Basis eines Datenmanagers Dokumente und Bilder hochgeladen und verwaltet werden können.

In der Praxis spielen diese Dienste im Hinblick auf ihre Reichweite im Vergleich zu Google. zum gegenwärtigen Zeitpunkt allerdings noch keine vergleichbare Rolle. Das liegt unter anderem daran, dass es sich bei der Vergabe von Tags um einen sehr

fehleranfälligen Prozess handelt. So kommt es unter anderem durch Rechtschreibfehler, fehlende Synonymerkennung (und dadurch verursachtes unkontrolliertes Wachstum der Menge der Tags) und Kombinationen unsachlich zugeordneter Tags nahezu zwangsläufig zu einer Verwässerung der Suchergebnisse. Es kann zwar etwas Ordnung in Tag-Sammlungen gebracht werden, indem den Benutzern Vorschläge beim Tagging unterbreitet werden. Fehlerhafte Zuordnungen werden hierdurch jedoch nicht verhindert. Als problematisch ist weiterhin die nur schwer zu gewährleistende Aktualität solcher Dienste anzusehen – z. B. bei der Nachrichtensuche –, da sie in der Regel auf eine längerfristige Speicherung von wichtigen Bookmarks ausgelegt sind. Aufgrund der kurzen Halbwertszeit von Nachrichten werden diese jedoch nur selten in die private Bookmarksammlung aufgenommen.

Dennoch liegen die Vorzüge, die ein Social Bookmarking-Dienst für den User schafft, klar auf der Hand: Die gespeicherten Bookmarks/Artikel/Dateien sind von jedem Rechner aus jederzeit erreichbar und darüber hinaus ggf. auch für die Allgemeinheit von Nutzen. Die Community dient zudem als Filter für die Bookmarks, um Inhalte mit geringerer Qualität auszusortieren. Thematisch können die gefundenen Treffer mehr in die Tiefe gehen, als bei algorithmischen Suchmaschinen. Auch können sich Gruppen einen Account mit Bookmarks teilen, die für ihre jeweiligen Mitglieder relevant sind (Vereine, Arbeitskollegen, Studenten).

3.2 Frage-Antwort-Dienste

Bei Frage-Antwort-Diensten handelt es sich ebenfalls um eine spezielle Ausprägungsform sozialer Softwarelösungen. Sie bieten dem Anwender die Möglichkeit, Fragen an andere Teilnehmer dieser Dienste zu stellen und beantworten zu lassen. Suchanfragen werden somit in einem von Menschen erzeugten Datenbestand durchgeführt, also von einer Community von Freiwilligen beantwortet. Dem liegt die Annahme zugrunde, dass algorithmenbasierte Suchmaschinen bestimmte Anfragen nicht ohne weiteres beantworten können. Exemplarisch hierfür seien z. B. Meinungsfragen oder Fragen nach Ratschlägen genannt. Als Anreiz zur Beantwortung von Fragen erhalten die Anwender in der Regel Punkte, die mit einem Rangsystem verknüpft sind (z. B. bei Lycos iQ). Je mehr Fragen ein Anwender beantwortet, desto höher kann er in seinem Rang aufsteigen und einen gewissen Status in der Community aufbauen. In technischer Hinsicht liegt den meisten Frage-Antwort-Diensten – analog den sozialen Bookmarksystemen – das Prinzip des Tagging zugrunde. Dies ermöglicht es z. B., Fragen in Themengebiete zusammenzufassen oder Anwender mit ähnlichen Interessen zu identifizieren, weist allerdings auch dieselben oben beschriebenen Schwachstellen auf (fehlerhafte Tags, unsachgemäße Zuordnungen usw.).

Dem Nutzer steht mittlerweile ein breites Angebot an Frage-Antwort-Diensten zur Verfügung, die sich in ihrer Funktionalität nicht wesentlich voneinander unterscheiden. Wichtiger als technische Features ist jedoch die Größe der Community. Nur wenn eine Frage in angemessener Zeit und zur Zufriedenheit des Fragestellers beantwortet wird, können solche Dienste einen Nutzen stiften. Je größer die Community dabei ausfällt, desto höher ist auch die Wahrscheinlichkeit, eine kompetente Antwort zu erhalten. In Deutschland spielen vor diesem Hintergrund

lediglich die Suchdienste Yahoo! Clever, Lycos iQ und guteFrage.net eine wichtige Rolle auf dem Markt.

Gemessen an ihrem wirtschaftlichen Erfolg kann gegenwärtig keiner der hier skizzierten sozialen Suchdienste überzeugen und einem Vergleich mit algorithmenbasierten Suchmaschinen standhalten. Auffallend ist, dass sich der Suchmaschinenriese Google aus diesem Geschäftsfeld komplett zurückgezogen hat; bis vor kurzer Zeit hat z. B. auch Google den kostenpflichtigen Frage-Antwort-Dienst Google Answers betrieben.

3.3 Semantische Suchdienste

Neben sozialen Suchdiensten werden weiterhin semantische Suchmaschinen als Alternative für algorithmenbasierte Suchmaschinen diskutiert. Um für deren Funktionsweise und Bedeutung genauer zu verstehen, ist es wichtig, zunächst auf die Vision des semantischen Webs im Allgemeinen einzugehen, die von Tim Berners-Lee geprägt wurde. Danach sollen Mensch und Computer ein gemeinsames Verständnis über einen bestimmten Ausschnitt der Realität haben und Webseiten inhaltlich interpretieren können. Zu diesem Zweck entwickelte Berners-Lee den Vorschlag, zusätzlich zu HTML, der Beschreibungssprache für die Ausgabe einer Webseite, einen Standard zur Beschreibung der Bedeutung der verwendeten Inhalte zu entwickeln und zu verwenden [13]. Diese Sprache würde es automatischen Verfahren und Computern erleichtern, die Inhalte einer Webseite zu verstehen und sinnvolle Schlussfolgerungen daraus zu ziehen.

Eine grundsätzliche Möglichkeit, die Bedeutung von Begriffen und deren Beziehungen untereinander zu modellieren sind Ontologien. Dabei handelt es sich um Datenstrukturen, in denen Begriffe als Konzepte und ihre Beziehungen als Relationen abgebildet werden. Würden Webseiten mit solcherart strukturierter Zusatzinformation versehen, wäre es z. B. möglich, mit automatischen Methoden zwischen gleichlautenden Begriffen mit unterschiedlicher Bedeutung zu unterscheiden (z. B. Golf als Auto, Sport oder Meeresbucht) oder Informationen zu ähnlichen Themen zu finden (z. B. von Golf zu Cricket als Sportart, die ebenfalls mit einem Schläger auf einem Rasen ausgeführt wird). Somit würden sich bei der Websuche viele Möglichkeiten eröffnen, Suchanfragen zu verstehen bzw. genauer zu formulieren.

Dabei stellt sich natürlich die Frage, wer diese Metainformationen zu den Webseiten erstellt und auf Korrektheit überprüft. Ontologien sind zwar ein mächtiges Hilfsmittel, stellen sich aber auch als Hindernis dar, wenn die Annotationen von durchschnittlichen Internetnutzern erstellt werden sollen. Hierzu sind Ontologien zu komplex und für den Laien nur schwer verständlich. Ein Ausweg liegt in der automatischen Erstellung von Ontologien. Dazu wird versucht, mit unterschiedlichen Methoden aus den Gebieten Text Mining, Informationsextrahierung und maschinelles Lernen, den Inhalt einer Seite automatisch zu erfassen, in Ontologiestrukturen abzulegen und mit bereits vorhandenen Ontologien zu verknüpfen. Allerdings kann man sich leicht vorstellen, dass auch dieser Ansatz sehr anspruchsvoll und daher noch nicht so ausgereift ist, dass es bereits darauf basierende verlässliche und robuste Produkte gäbe.

Insgesamt muss man also feststellen, dass das semantische Web als zusätzliche Bedeutungsschicht, die das World Wide Web vollständig umfasst, eine Vision ist, die vermutlich so nicht erreichbar sein wird. Ein Beispiel dafür ist der in der Presse öfters

als „Google-Killer“ gehandelte semantische Suchdienst Powerset, der trotz einer relativ langen Entwicklungsphase noch immer nicht wie geplant online ist und sich derzeit noch als Wikipedia-Suchmaschine beschränkt; und auch hier sind die Ergebnisse noch nicht zufrieden stellend.

Dennoch ist es beachtlich, welche Fortschritte in der Forschung bereits erzielt wurden. Zieht man dazu noch das dem Normalnutzer zugänglichere Tagging zur Erzeugung von Metainformationen in Betracht, so lassen sich Vorgehensweisen vorstellen, die aus den zwar unstrukturierten und fehlerbehafteten, aber von menschlichen Benutzern bereitwillig erstellten Zusatzinformationen in Kombination mit bereits vorhanden Ontologien und semantischen Verfahren die gewünschten Annotationen erzeugen.

Eine andere Möglichkeit zur Verbreitung semantischer Annotationen, die sich derzeit immer größerer Verbreitung erfreut, ist die Verwendung so genannter Mikroformate im HTML-Code der Webseite. Hierdurch bekommen die Entwickler von Webseiten die Möglichkeit, bestimmte Informationen in Standardformaten direkt als unsichtbare Informationen auf der Webseite zu hinterlegen. Sobald solche Mikroformate von den großen Suchmaschinenbetreibern unterstützt werden (wie kürzlich von Yahoo! Angekündigt [14]) und dadurch die Aussicht besteht, dass die gesuchte Information von den Standardsuchmaschinen besser gefunden wird, werden weitere Seitenbetreiber Mikroformate einsetzen. Eine breite Toolunterstützung, die das Erstellen von Mikroformaten erleichtert oder gar automatisiert, würde wohl bald folgen. Ist so ein Punkt erreicht, könnte zumindest ein Teil der Semantic-Web-Vision Realität werden.

3.4 Konvergente Suchdienste

Neben den bislang skizzierten Suchdiensten existieren noch weitere, die Eigenschaften mehrerer Suchdienste vereinen und als konvergent bezeichnet werden (vgl. Abb. 1.1). Sie verfolgen generell das Ziel, die Vorteile der bislang nur isoliert voneinander fungierenden Suchdienste miteinander zu verknüpfen, um die Qualität der Suchergebnisse zu verbessern oder zumindest deren Nachteile zu mindern. Ein Beispiel hierzu ist die Suchmaschine Wikia Search, die vom Wikipedia-Gründer Jimmy Wales frei zur Verfügung gestellt wird und das Wikipedia-Prinzip des freien Editierens mit den Eigenschaften einer algorithmischen Suchmaschine verbinden soll. Das bedeutet, dass mittels Crawling ein Index erstellt wird, der dann von den Nutzern bearbeitet werden soll. Die Suchergebnisse des im Januar 2008 eingeführten Dienstes sind wie erwartet noch nicht vergleichbar mit anderen Suchmaschinen und erst die nächste Zeit wird zeigen, ob aus diesem Dienst ein ähnlich erfolgreiches Projekt wie Wikipedia entstehen wird. Weitere Beispiele für die konvergente Suche sind die personalisierte Suche und die implizite Suche in so genannten Clickstreams. So soll der derzeit in der Betaphase befindliche Suchdienst Twine. Komponenten sozialer Netzwerke und semantischer Suche in einer persönlich konfigurierbaren Suche zusammenführen [15, 16]. Basierend auf einer Semantic-Web-Datenstruktur, die es ermöglicht, Inhalte computerverständlich abzulegen, soll der Nutzer dabei unterstützt werden, seine gefundenen Informationen mit anderen auszutauschen und selber wiederum durch den Austausch mit anderen besser an gesuchte Informationen zu gelangen. Bei Clickstreams hingegen werden die impliziten Informationen ausgenutzt, die der Nutzer durch sein Surfverhalten hinterlässt. Klicks auf Links und

Seitenverweildauern werden mit Methoden des Data Mining und des maschinellen Lernens ausgewertet, um daraus zu folgern, was der Nutzer „eigentlich“ suchte, ohne eine explizite Suchanfrage gestellt zu haben. Da solche Verfahren allerdings nur mit Hilfe der Auswertung persönlicher Daten auf den Systemen der Internetprovider funktionieren können, stößt dieser Ansatz in der öffentlichen Diskussion auf breite Kritik.

4. FAZIT

Jeder der in diesem Beitrag untersuchten Suchdienste birgt Vor- und Nachteile, die für eine Empfehlung, welcher Dienst bei welcher Art von Suche herangezogen werden sollte oder wie die Ansätze ggf. zu kombinieren sind, miteinbezogen werden müssen. Die algorithmische Suchmaschine bietet dem Nutzer mit aktuellen Meldungen und News eine Aktualität, die andere Dienste in diesem Umfang nicht bieten. Eine Suche nach variablen Datenformaten und verschiedenen Kategorien wie Nachrichten, Karten, Shops und Bildern leistet in vollem Umfang ebenfalls nur die algorithmische Suche. Weder Frage-Antwort- noch Social Bookmarking-Dienste generieren Suchergebnisse in dieser Menge unter Abdeckung aller Themenbereiche. Diese große Menge der Suchergebnisse ist aber gleichzeitig ein Nachteil, da nicht alle Treffer für die Suchanfrage relevant sind. Ebenso sind hier das Problem des Suchmaschinen-Spammings und die Schwierigkeiten durch die Uneindeutigkeit menschlicher Sprache zu bedenken. Der Nutzer muss also durch eine große Anzahl von Treffern bewerten, um das für ihn relevante Ergebnis zu finden. Dabei enthält der Index einer algorithmischen Suchmaschine noch nicht einmal ansatzweise alle im WWW enthaltenen Webseiten und Informationen.

Eine mögliche Lösung für diese Probleme könnte in konvergenten Diensten liegen – einer Kombination einer algorithmischen Vorgehensweise mit einem Social Bookmarking-Dienst, da über eine Anfrage durch einen Social Bookmarking-Dienst auch speziellere Webseiten gefunden und durch die Community bewertet werden können. Somit bieten Social Bookmarking-Dienste die kollektive Intelligenz der Community zur Steigerung der Qualität der Suchergebnisse. Nachteilig beim Social Bookmarking-Dienst sind wiederum die fehlende Aktualität und eventuelle thematische Lücken, da hier ja nicht automatisiert aktualisiert und indexiert wird. Dennoch ist Social Bookmarking eine interessante Variante, tiefer ins Web einzutauchen.

Durch die Kombination mit einem Frage-Antwort-Dienst könnte dem User ein zusätzlicher Service geboten werden. Die Stärke der Frage-Antwort-Dienste liegt in der Beantwortung von gezielt gestellten individuellen Fragen. Fragen im Erfahrungsbereich können weder von einer algorithmischen Suchmaschine noch von einem Social Bookmarking-Dienst beantwortet werden. Die Richtigkeit der Antwort kann ebenfalls nicht garantiert werden. Auch wird es hierbei immer Lücken in den behandelten Themenbereichen geben, da nur Antworten auf Fragen gegeben werden können, die von Nutzern gestellt wurden. Unter Umständen muss der Nutzer dann eine längere Zeit auf die Beantwortung seiner Frage warten. Durch den starken Community-Charakter der Frage-Antwort-Dienste liegt der Schwerpunkt weniger auf wissenschaftlichen Themen, sondern eher auf Themengebieten wie Freizeit, Kultur

und Medien. Dadurch haben Frage-Antwort-Dienste oft eher einen Chat- oder Forencharakter.

Zusammenfassend betrachtet kann man jedoch feststellen, dass sich trotz aller offensichtlichen Herausforderungen schon bei dieser ersten Betrachtung der existierenden Suchdienste und ihrer prinzipiellen Kombinationsmöglichkeiten ein Bild ergibt, das den Versuch einer Integration insbesondere sozialer Suchansätze in algorithmische Vorgehensweisen als lohnenswert erscheinen lassen. Es soll daher in diesem vorliegenden Band detailliert untersucht werden, wie die betrachteten Dienste genau funktionieren, welche Erfahrungen damit gemacht wurden, welche Qualitätsunterschiede sich messen lassen, und ob eine Kombination tatsächlich sinnvoll scheint.

Verwendete Literatur

1. Gasser, Urs/Thurman, James (2007): Themen und Herausforderungen der Regulierung von Suchmaschinen. In: Machill, Marcel/Beiler, Markus (Hrsg.): Die Macht der Suchmaschinen, Köln 2007, S. 44-64.
2. Die Grünen (2005): Suchmaschinen: Das Tor zum Netz. Onlinedokument: <http://www.gruene-bundestag.de/cms/publikationen/dokbin/63/63265.pdf> [Abruf am 14.05.2008]
3. Die Grünen (2006): Google verleibt sich YouTube ein – und wird noch mächtiger. Onlinedokument: http://www.gruene-bundestag.de/cms/presse/dok/150/150791.google_verleibt_sich_youtube_ein_und_wir.html [Abruf am 14.05.2008]
4. Gray, Matthew (1996): Internet Statistics. Growth and Usage of the Web and the Internet. Onlinedokument: <http://www.mit.edu/~mkgray/net/> [Abruf am 24.05.2008]
5. ComScore (2008): Suchabfragen im April 2008. Onlinedokument: www.comscore.com [Abruf am 01.06.2008]
6. Schäfer, Karim (2006): Medienbranche mit guten Perspektiven bis 2010: Digitalisierung schafft robustes Wachstum. In: Innovations Report, 21.06.2006.. Onlinedokument: http://www.innovations-report.de/html/berichte/kommunikation_medien/bericht-66578.html [Abruf am 14.05.2008]
7. Microsoft Advertising (2008): Das Internet entwickelt sich bis 2010 zum weltweit drittgrößten Medium. Onlinedokument: http://advertising.microsoft.com/deutschland/Branchen-News?Adv_NewsID=891 [Anruf am 14.05.2008]
8. OVK (2007): Online-Report 2007/02. Onlinedokument: http://www.bvdw.org/fileadmin/downloads/marktzahlen/basispraesentationen/OVK_OnlineReport_2007_02.pdf [Abruf am 14. 05 2008]
9. Welt Online (2007): Google knackt die Milliarden-Marke. In: Welt Online, 18.10.2007. Onlinedokument: http://www.welt.de/wirtschaft/article1278520/Google_steigert_Gewinn_um_46_Prozent.html [Abruf am 24.05.2008]
10. Komus, Ayelt (2006): Social Software als organisatorisches Phänomen – Einsatzmöglichkeiten im Unternehmen. In: Hildebrand, Knut/Hofman, Josephine (Hrsg.): Social Software. HMD – Praxis der Wirtschaftsinformatik, H. 252, S. 36-14.
11. Gräfe, Gernot/Maaß, Christian./Heß, Andreas (2007): Alternative Search Engines: Seven Hypotheses Regarding the Future of Social Bookmarking. In: Software, Agents and Services for Business, Research and E-Sciences - Conference on Social Semantic Web (SABRE/CSSW 2007), Leipzig 2007.

12 Status quo und Entwicklungstendenzen

12. Bächle, Michael (2006): Social Software. In: Informatik Spektrum Bd. 29, H. 2, S. 121-124.
13. Berners-Lee, Tim/ Hendler, James/ Lassila, Ora (2001). The Semantic Web. In: Scientific American Magazine, Mai 2001. Onlinedokument: <http://www.sciam.com/article.cfm?id=the-semantic-web&ref=sciam> [Abgerufen am 18.03.2008]
14. Kumar, Amit (2008): The Yahoo! Search Open Ecosystem. In: Yahoo! Search Blog, 13.03.2008. Onlinedokument: <http://www.ysearchblog.com/archives/000527.html> [Abruf am 18. 03. 2008]
15. MacManus, Richard (2007): Twine: The First Mainstream Semantic Web App? In: ReadWriteWeb, 18.10.2007. Onlinedokument, http://www.readriteweb.com/archives/twine_first_mainstream_semantic_web_app.php [Abruf am 26.03.2008]
16. Schonfeld, Erick (2007): Twine Launches A Smarter Way To Organize Your Online Life. In: TechCrunch, 19.10.2007. Onlinedokument: <http://www.techcrunch.com/2007/10/19/twine-launches-a-smarter-way-to-organize-your-online-life/> [Abruf am 26.03.2008]

Qualitative Laboruntersuchung zur Anwendung von Social Software-Systemen

ANNELIESE BERGEN
anneliese.bergen@haw-hamburg.de

TINA KRAUSE
Krause-Tina@t-online.de

Ziel dieser Arbeit ist die Untersuchung der Bekanntheit und Anwendung von Social Software auch im Hinblick auf Begriffsvorstellungen, der Herangehensweise bei Suche und Tagging mit Hilfe von Frage-Antwort- und Social Bookmarking-Diensten durch den Nutzer sowie der Zufriedenheit mit Suchergebnissen. Hierfür wurde eine qualitative Laboruntersuchung mit Fragebogenerhebung durchgeführt.

Das Ergebnis zeigt, dass die Bekanntheit von Frage-Antwort-Diensten beschränkt ist; die Bedeutung der Begriffe Social Bookmarking und Tagging ist der Mehrzahl ebenfalls nicht bekannt. Die Herangehensweise bei der Suche mit den untersuchten Diensten ist von Unsicherheit gegenüber der Funktionsweise und richtigen Bedienung geprägt. Auch die Zufriedenheit mit den Ergebnissen ist begrenzt. Trotz der geringen begrifflichen Kenntnis des Tagging wurde diese Funktion zumeist richtig erkannt und eingesetzt, wobei eher allgemeine Tags gewählt wurden.

Vorschläge zur Verbesserung der untersuchten Dienste sind die Einschränkung von Selektionsmöglichkeiten, die Anpassung der Suchfunktionen an gewohnte Anwendungen wie Google und das Angebot einer relevanzsortierten Ergebnisseite. Eine Einbindung von Suchergebnissen alternativer Suchdienste in Trefferseiten von Suchmaschinen wird nur bei fehlender Erschließung der Internetquellen durch Suchmaschinen-Indizes empfohlen.

1. EINLEITUNG²

Im Zeitalter des Web 2.0 dienen nicht nur Suchmaschinen dem Internetnutzer zur Informationssuche, auch Social Software-Anwendungen sollen neue Wege der Informationssuche aufzeigen. Die Anforderungen an solche Anwendungen sind die gleichen wie bei Suchmaschinen: Sie müssen einfach und schnell zu bedienen sein, damit ein möglichst großer Kreis von Internetusern diese bedienen und nutzen kann. Doch können diese neuen Internetdienste mit den traditionellen Suchmaschinen mithalten? Sind Internetnutzern die Begriffe Frage-Antwort-Dienst, Social Bookmarking und Tagging geläufig?

Das Ziel der im Folgenden beschriebenen Untersuchung war es herauszufinden, welche dieser Dienste und ihrer jeweiligen Features die Nutzer zur Informationssuche kennen und anwenden. Insbesondere stellte sich die Frage, wie die Nutzer an ihre

² Für die Unterstützung zur Durchführung der Laboruntersuchung bedanken wir uns bei unserer Kommilitonin Meike Paustian.

Suche herangehen und wie zufrieden sie mit den erhaltenen Suchergebnissen sind. Schließlich sollte analysiert werden, ob den Nutzern der Begriff Tagging bekannt ist und wie bei der Vergabe von Tags vorgegangen wird. Ein weiteres Ziel bestand darin, Vorschläge für die Einbindung der Ergebnisse aus Frage-Antwort-Diensten und Social Bookmarking in die Suchergebnisse von Suchmaschinen zu entwickeln. Um diese Ziele zu erreichen, wurde eine qualitative Laboruntersuchung durchgeführt. In dieser Untersuchung wurden Testpersonen bei der Verwendung der genannten Social Software-Anwendungen getestet. Die dabei gewonnenen Erkenntnisse sollten Hinweise zur Optimierung der Angebote aufzeigen, die bei der Informationssuche mit den genannten Diensten von den Testpersonen ermittelt worden sind.

Im Anschluss daran wurde aus den Erkenntnissen ein Fazit gezogen, das Vorschläge für eine Integration der Suchergebnisse aus Frage-Antwort-Diensten und Social Bookmarking in die Ergebnisse von Suchmaschinen enthält.

2. QUALITATIVE LABORUNTERSUCHUNG

Die qualitative Laboruntersuchung wurde im Usability Labor der HAW Hamburg durchgeführt. Im Unterschied zu einem herkömmlichen Usability Test war es nicht das Ziel der Laboruntersuchung, eine Website mit Hilfe verschiedener Testpersonen auf ihre Gebrauchstauglichkeit zu untersuchen. Die Erkenntnisse sollten nicht dazu dienen, Hinweise für eine Verbesserung einer Website zu liefern.

Die qualitative Laboruntersuchung sollte vielmehr aufzeigen, wie Internetnutzer in einer bestimmten Situation an die Informationssuche herangehen. Ein besonderes Interesse bestand darin herauszufinden, wie die Rechercheergebnisse aus den Frage-Antwort- bzw. den Social Bookmarking-Diensten beurteilt werden. Des Weiteren sollte analysiert werden, wie die Nutzer beim Taggen vorgehen und ob ihnen die Funktion von Tags bei der Informationssuche vertraut ist.

2.1 Technischer Ablauf

Wie auf Abbildung 2.1 zu sehen ist, befindet sich im Usability Labor ein Tisch mit Stühlen. An diesem Tisch findet vor Beginn der Untersuchung das erste Gespräch mit der Testperson statt, in welchem der Ablauf der Laboruntersuchung erläutert wird. Des Weiteren befindet sich im vorderen Laborteil der technische Arbeitsbereich mit einem Arbeitsplatz (1), der die Cursor-Bewegungen aufzeichnet. Ebenfalls befinden sich hier der Kontrollmonitor (3) sowie der DVD-Rekorder und das Mischpult für die Kamera- und Toneinstellungen (2).

Dieser technische Bereich wird während der Laboruntersuchung mit Stellwänden vom Arbeitsbereich der Testperson abgetrennt, damit diese so wenig wie möglich durch die technische Aufzeichnung abgelenkt wird.

Hinter der Abtrennung befindet sich der Arbeitsbereich der Testperson mit Computer, Monitor, Tastatur und Maus (4). Auf dem Monitor ist eine Webcam angebracht, die die Testperson während der Untersuchung filmt. Des Weiteren wird das Profil der Testperson von der Kamera 1 in Nahaufnahme aufgezeichnet. Die Kamera 2 filmt diese zusammen mit dem Moderator. Auf dem Tisch befinden sich zwei Mikrofone (M1, M2), die das Gespräch zur Aufzeichnung an den technischen Arbeitsbereich weiterleiten.

Durch die Mikrofone, die Bildschirmaufzeichnung und die verschiedenen Kameraperspektiven soll die Testsituation für die spätere Auswertung möglichst genau aufgezeichnet werden. Die Kamera am Bildschirm filmt beispielsweise nur das Gesicht der Testperson. Dadurch sollen Gesichtsausdrücke aufgezeichnet werden, die Aufschluss über Reaktionen der Testperson geben, auch wenn diese nicht mündlich geäußert werden. Ein fragender, überraschter oder unzufriedener Gesichtsausdruck kann so beispielsweise in Beziehung zu einer bestimmten Situation gesetzt und interpretiert werden.

Auch die Bildschirmanzeigen werden aufgezeichnet, wobei der Cursor optisch hervorgehoben wird. So kann für die Auswertung nachvollzogen werden, welche Bereiche der Website die Testperson mit der Maus angesteuert hat, welche Links die Testperson angeklickt hat und welche Inhalte in Such- und Eingabefelder auf der Website eingegeben wurden.

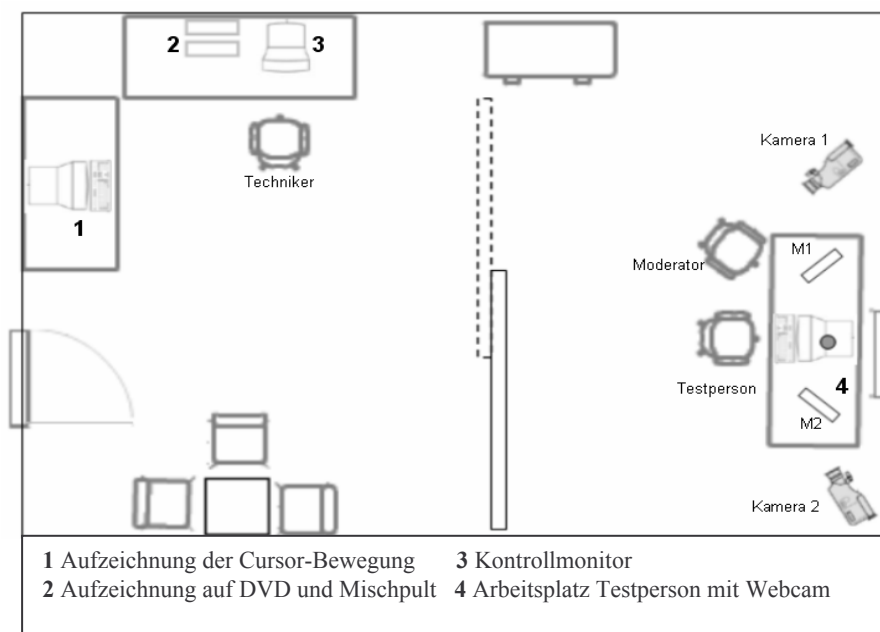


Abb. 2.1 Aufbau des Usability Labors der HAW Hamburg

2.2 Durchführung

Für die qualitative Laboruntersuchung mussten zunächst Testpersonen ausgewählt und Termine vereinbart werden. Bei der Auswahl der Testpersonen wurden keine besonderen Vorgaben gemacht. Es wurde lediglich darauf geachtet, dass die Testpersonen möglichst unterschiedlichen Altersgruppen angehören und dass deren Vorkenntnisse, bezogen auf die Nutzung des Internets, möglichst unterschiedlich ausgeprägt sind (vgl. Anhang 1). Somit setzte sich die Gruppe aus sechs männlichen und vier weiblichen Testpersonen zusammen. Diese sind zum größten Teil zwischen 20 und 40 Jahren alt und gehören verschiedenen Berufsgruppen an (vgl. Anhang 2).

Bei der Durchführung der qualitativen Laboruntersuchung waren jeweils eine Testperson und zwei Projektteilnehmer anwesend. Dabei hatte ein Projektteilnehmer die Aufgabe, die technischen Einstellungen vorzunehmen und die Bild- und Tonaufzeichnungen zu überwachen. Während der Untersuchung sollten gleichzeitig die Fragen der Testperson notiert werden, um diese nach Ablauf der Untersuchung beantworten zu können. Der zweite Projektteilnehmer war dafür zuständig, die Laboruntersuchung zu moderieren. Dazu gehörte unter anderem, der Testperson Aufgaben zu stellen und diese zum „Lauten Denken“ zu animieren.

„Im Rahmen von Usability-Tests hat sich die so genannte Thinking-aloud-Methode bewährt. Die Testpersonen führen eine konkrete Aufgabe durch und werden gebeten, dabei laut zu denken.“ [1]. Dadurch sollen „Rückschlüsse auf ihre Denk- und Entscheidungsprozesse“ (vgl. [2], S. 212) ermöglicht werden. Diese wertvollen Informationen können zusätzlich zu den nonverbalen Reaktionen der Testpersonen ausgewertet und gegebenenfalls mit diesen verknüpft werden. Diese Methode wurde bei der Laboruntersuchung angewandt, um möglichst viele Gedanken, Anmerkungen und Fragen der Testperson aufzeichnen und auswerten zu können.

Ein für die qualitative Laboruntersuchung eigens konzipierter Leitfaden diente dem Moderator als Hilfsmittel für die Durchführung der Laboruntersuchung (vgl. Anhang 3). Dadurch sollten alle Untersuchungen, auch von verschiedenen Moderatoren, möglichst ähnlich und somit vergleichbar durchgeführt werden.

Zur Durchführung der Laboruntersuchung wurde im Vorfeld für jede Testperson ein eigener Nutzeraccount bei Lycos iQ angelegt. Mit diesem Account wurde die Testperson noch vor Beginn der Untersuchung eingeloggt. Damit sollte gewährleistet werden, dass der Testperson während der Durchführung der Laboruntersuchung alle Funktionen zur Verfügung stehen. Bei drei der Testdurchführungen war ein Einloggen in den Nutzeraccount nicht möglich. Ebenso wurde bei Mister Wong ein Testaccount für die Testpersonen angelegt.

Vor Beginn der Laboruntersuchung wurde die jeweilige Testperson im Foyer der HAW Hamburg abgeholt und ins Usability Labor begleitet. Dort wurden beide Projektteilnehmer vorgestellt. Nach einem Dank für die Teilnahme an der Untersuchung wurde der Ablauf erläutert und ggf. wurden Fragen der Testperson beantwortet. Um während der Untersuchung Film- und Tonaufnahmen aufzeichnen und diese für die Auswertung verwenden zu dürfen, wurde die Testperson um ihr schriftliches Einverständnis gebeten (vgl. Anhang 4).

Zu Beginn der Laboruntersuchung wurden die Testpersonen aufgefordert, sich die Website Lycos iQ anzuschauen und ihre ersten Eindrücke bezüglich des Erscheinungsbildes sowie der erkennbaren Funktionen zu äußern. Somit sollte ermittelt werden, was die Personen positiv anspricht oder auch negativ berührt. Des Weiteren diente dieser Punkt zur Hinterfragung, welche Funktionen den Testpersonen verständlich sind, bzw. welche Erwartungen diese an bestimmte Funktionen haben.

Die wesentlichen Punkte der Testdurchführung waren drei verschiedene Situationsaufgaben, die der Testperson gestellt wurden. Hierbei wurden die Aufgaben der Person durch den Moderator mündlich erläutert und zur Unterstützung auch in ausgedruckter Form vorgelegt.

Die Testpersonen sollten sich in die folgenden Situationen hineinversetzen:

- Situation 1: Meinungs- und Empfehlungsfragen (vgl. Anhang 5)
Die Testperson sollte sich vorstellen, dass sie ein Produkt erwerben möchte und sollte für dieses Produkt Meinungen und Empfehlungen von anderen Lycos iQ-Nutzern in Erfahrung bringen. Die Situationen waren so ausgearbeitet, dass es den Testpersonen möglich gewesen wäre, diese Informationen durch Recherche in den Fragen und Antworten von Lycos iQ in Erfahrung zu bringen.
Bei dieser Aufgabe sollte ermittelt werden, mit welchen Suchstrategien die Testpersonen vorgehen und wie die erhaltenen Ergebnisse bewertet werden. Da es sich hier um Meinungs- und Empfehlungsinformationen handelte, gab es keine vordefinierte richtige oder falsche Antwort.
- Situation 2: Faktenfrage (vgl. Anhang 5)
Die Testperson wurde aufgefordert, eine ganz bestimmte Information in Erfahrung zu bringen. Da es sich bei dieser Aufgabe um die Ermittlung eines bestimmten Faktums handelte, kam auch nur eine richtige Antwort in Frage. Im Vorfeld wurde sichergestellt, dass sich diese Information nicht mittels Recherche in Lycos iQ finden lässt.
Hierbei sollte gezielt ermittelt werden, wie die Testpersonen bei der Einstellung einer eigenen Frage vorgehen. Somit konnten die einzelnen Schritte der Frageeinstellung analysiert werden.
- Situation 3: Bookmarks und Links (vgl. Anhang 5)
Die Testperson sollte in dieser Situation gezielt nach Websites suchen. Hierbei sollte die Herangehensweise der Person ermittelt werden. Ebenso stellte sich die Frage, ob die Testpersonen Bookmarks erkennen und wie sie diese beurteilen.
Diese Situation hatten alle Testpersonen auf der Website Lycos iQ zu ermitteln.
Zusätzlich wurden die Testpersonen aufgefordert, die gleiche Situation auf der Website Mister Wong zu lösen. Somit sollten Vergleiche zwischen diesen beiden Websites zur Handhabung von Bookmarks und Tags gezogen werden. Aus zeitlichen Gründen wurde diese Aufgabe nur drei Testpersonen auf der Website Mister Wong gestellt.

Nach Beendigung der Situationsaufgaben wurde ein kurzes Interview mit der Testperson geführt. Hierbei sollte der Gesamteindruck bezüglich der Website und einzelner Funktionen in Erfahrung gebracht werden. Im Vordergrund stand hierbei die Hinterfragung der in den Situationsaufgaben genutzten Situationen. Abschließend wurde die Testperson befragt, ob sie sich eine weitere Nutzung der Website Lycos iQ vorstellen könnte, bzw. welche Funktionen künftig genutzt werden würden.

Schließlich wurde die Testperson gebeten, einen kurzen Fragebogen (vgl. Anhang 1) auszufüllen. Mit Hilfe dieses Fragebogens sollten statistische Angaben ermittelt werden, wie z.B. Alter, Geschlecht und Beruf der Testperson. Weiterhin sollten verschiedene Angaben über die Vorkenntnisse zum Thema der Untersuchung gemacht werden. Dazu gehörten im Allgemeinen die Nutzung des Internets und im Besonderen die Nutzung von Suchmaschinen, Frage-Antwort-Diensten, Social Bookmarking und Tags. Aus den Ergebnissen dieser Befragung (vgl. Anhang 2) sollten eventuelle Zusammenhänge zwischen den Vorkenntnissen der Testperson und den Erkenntnissen aus der jeweiligen Untersuchung ermittelt werden.

Nach der Durchführung der Laboruntersuchung wurden der Testperson ihre noch offenen Fragen beantwortet. Im Anschluss wurde die Testperson darum gebeten, zwei weitere kurze Feedbackbögen auszufüllen. Mit einem dieser Feedbackbögen (vgl. Anhang 6) sollte die Zufriedenheit mit der Website insgesamt sowie mit einzelnen Funktionen erfragt werden. Zusätzlich sollte die Testperson Auskunft darüber geben, ob und für welche Zwecke eine zukünftige Nutzung der Website in Betracht gezogen wird. Die Auswertung befindet sich im Anhang 7.

Der zweite Feedbackbogen (vgl. Anhang 8) bezog sich auf die Durchführung der Untersuchung durch die Projektteilnehmer. Durch dieses Feedback sollten mögliche Vorschläge zur Verbesserung der Durchführung ermittelt werden.

2.3 Auswertung der Laboruntersuchung

Um aus der Laboruntersuchung Erkenntnisse herleiten zu können, wurde ein tabellarisches Schema zur Auswertung erstellt. Mit Hilfe dieses Schemas sollte ein Vergleich der Vorgehensweise der verschiedenen Testpersonen erleichtert werden. Darin wurde unter anderem erfasst, wie die Testpersonen an die gestellten Aufgaben herangegangen sind und wie sie ihre Ergebnisse beurteilt haben. Das Schema wurde mit Hilfe des Tabellenkalkulationsprogramms Microsoft Excel erstellt. Dadurch sollten die Daten möglichst übersichtlich dargestellt und teilweise auch durch graphische Darstellungen veranschaulicht werden.

Zur Auswertung der Laboruntersuchung wurden verschiedene Aufzeichnungen herangezogen. Dies waren zum einen die Film- und Tonaufnahmen, die während der Untersuchung aufgezeichnet und auf DVDs gespeichert wurden. Zum anderen wurde die Bewegung des Cursors auf dem Bildschirm aufgezeichnet, um zu ermitteln, wie sich die Testpersonen auf der Website bewegen und welche Eingaben sie in die Such- und Eingabefelder machen.

Zusätzlich wurden während der Untersuchung Fragen der Testperson sowie besondere Situationen und Äußerungen notiert. Dadurch sollte sichergestellt werden, dass den Testpersonen noch offene Fragen zu den Funktionen der Website oder unklare Begriffe nach der Untersuchung erläutert werden konnten, um die Zufriedenheit der Testpersonen zu erhöhen. Schließlich wurden zur Auswertung der Laboruntersuchung die Ergebnisse einer kurzen schriftlichen Befragung herangezogen (vgl. Anhang 2).

2.4 Suche allgemein

Auf der Lycos iQ-Startseite (vgl. Abb. 2.2) werden den Nutzern verschiedene Funktionen und Möglichkeiten geboten, um nach Informationen zu recherchieren. Besonders auffällig für die Testpersonen sind hier das obere Suchfeld sowie das mittlere Fragefeld. Einem Teil der Testpersonen war anfangs nicht klar, wo der Funktionsunterschied dieser beider Felder liegt. Intuitiv entschieden sich die Testpersonen aber größtenteils ihre Suchanfragen in das obere Suchfeld einzugeben, da dieses ihnen vertraut erschien.

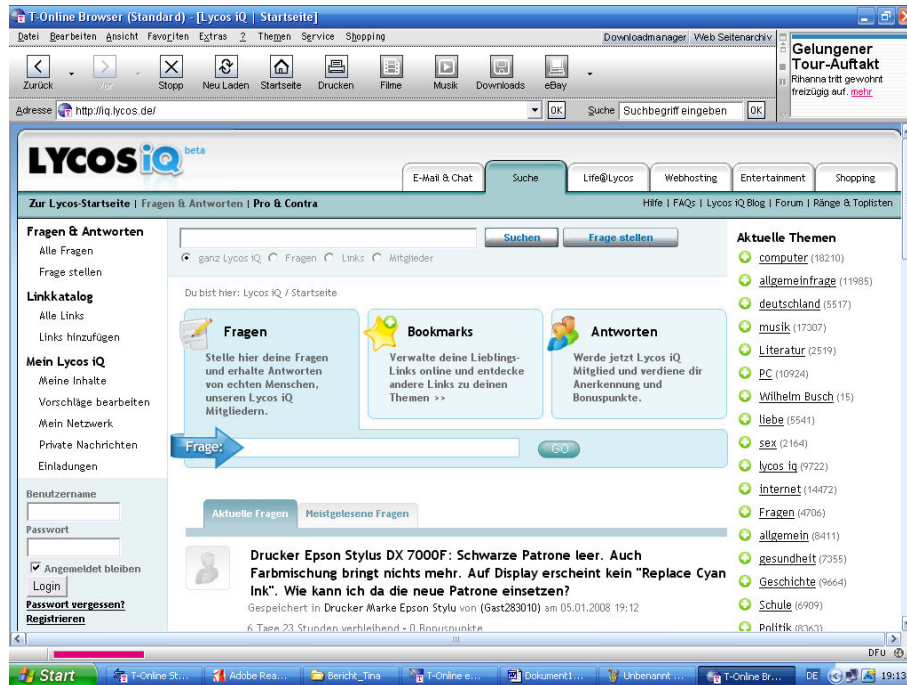


Abb. 2.2 Screenshot der Lycos iQ-Startseite

Insgesamt wurden während der gesamten Laboruntersuchung 56 Suchanfragen gestellt. Dabei wurde durchschnittlich mit 2,16 Stichwörtern gesucht. Bei 18 Prozent aller Suchanfragen wurde den Suchbegriffen ein „+“ vorangestellt. Die Testpersonen gaben auf Nachfrage an, dass ihnen die Verknüpfung eines Suchbegriffs mit dem „+“ aus der Suche mit Suchmaschinen (überwiegend Google) bekannt ist.

Die Selektionsmöglichkeit unterhalb des Suchfeldes (vgl. Abb. 2.19), mit der ausgewählt werden kann, in welchem Bereich der Lycos iQ-Website gesucht werden soll, wurde von den Testpersonen zu 59 Prozent bewusst gewählt. Überwiegend haben die Testpersonen die Option „ganz Lycos“ ausgewählt (vgl. Abb. 2.3).

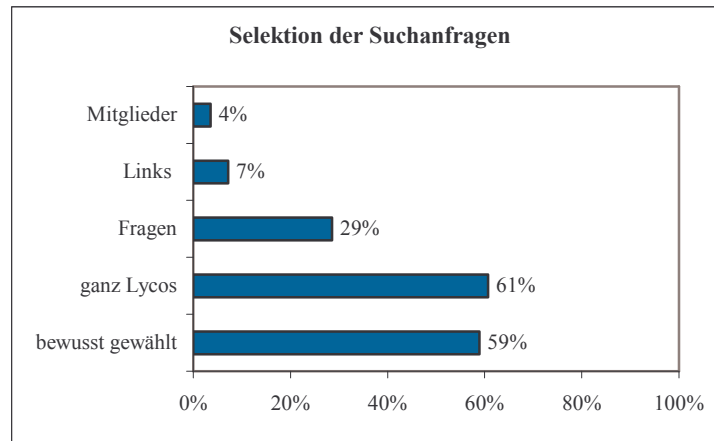
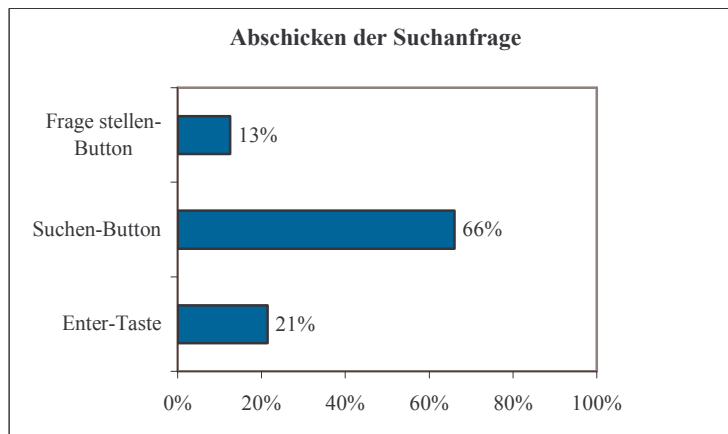


Abb. 2.3 Selektion der Suchanfragen (n= 56)

Unter den Testpersonen kamen Missverständnisse bei der Selektionsmöglichkeit „Fragen“ auf. Hier war diesen anfangs nicht bewusst, dass mit dieser Selektion nicht zwangsläufig die Einstellung einer eigenen Frage verbunden ist. Ebenso kam seitens einiger Testpersonen die Vermutung auf, dass nur in den Fragestellungen und nicht in den dazugehörigen Antworten gesucht wird. Die Option „Mitglieder“ war den Testpersonen nicht verständlich, da sie keinen Sinn darin gesehen haben, in diesen zu suchen. Es wurde vermutet, dass diese Funktion dazu dient, bekannte Mitglieder zu finden.

Abb. 2.4 Möglichkeiten zum Abschicken einer Suchanfrage
(n=56)

Bei der Laboruntersuchung hat sich gezeigt, dass die Suche bewusst mit dem Button zum Starten der Suche bzw. zum Einstellen einer Frage abgeschickt wurde (vgl. Abb. 2.4). Lediglich 21 Prozent aller Suchanfragen sind mit dem Betätigen der Enter-Taste abgeschickt worden. Hierbei geschah das Abschicken der Suchanfrage intuitiv. Daher war eine Testperson sehr verwundert, dass sie direkt zur Frageeinstellung weitergeleitet wurde und keine Rechercheergebnisse angezeigt

bekommen hat. Diese Testperson hatte nicht bemerkt, dass die Option „Fragen“ unterhalb des Suchfeldes selektiert war und sie somit durch Betätigen der Enter-Taste direkt zur Frageeinstellung geführt worden ist.

Über die Hälfte aller in der Laboruntersuchung durchgeführten Suchanfragen (55 Prozent) wurden noch einmal mit veränderten oder erweiterten Suchbegriffen durchgeführt. Gründe hierfür waren die noch nicht ausreichende Zufriedenheit mit den Suchergebnissen bzw. die Absicherung, ob sich mit anderen Suchbegriffen nicht mehr Ergebnisse erzielen lassen.

2.5 Situationsaufgabe 1: Meinungs- und Empfehlungsfragen

Die erste Aufgabe (vgl. Anhang 5) haben acht Testpersonen mit einer Recherche begonnen. Hierbei wurden die Suchbegriffe in das obere Suchfeld eingegeben. Bei der Auswahl des zu durchsuchenden Bereichs von Lycos iQ entschieden sich sechs Personen für „ganz Lycos“ und zwei Personen für den Bereich „Fragen“.

Die sofortige Einstellung einer Frage bei den verbliebenen zwei Testpersonen ist damit zu begründen, dass diese ihre Suchbegriffe direkt in das Fragenfeld eingegeben haben und somit zur Frageeinstellung weitergeleitet wurden. Somit haben diese eine eigene Frage eingestellt und die Möglichkeit einer eigenen Recherche nicht weiter verfolgt.

Bei dieser Aufgabenstellung haben neun Testpersonen ein für die Situation relevantes Rechercheergebnis erwartet. Als Rechercheergebnis wurde hier die Anzeige einer Linkliste erwartet, wie es von der Suchmaschine Google her bekannt ist. Lediglich eine Testperson hat Suchergebnisse aus den auf Lycos iQ vorhandenen Fragen und Antworten erwartet.

Bei der Anzeige der Treffer ist den Testpersonen aufgefallen, dass die gefundenen Treffer aus den verschiedenen Bereichen (Fragen und Antworten, Links, Mitglieder) stammen. Unklar war den Testpersonen auch hier wieder die Rubrik der Mitglieder, da kein Bezug zur eigenen Suchanfrage hergestellt werden konnte. Auffällig war ebenfalls, dass sich die Testpersonen größtenteils nur die erste Ergebnisseite angesehen und gescrollt haben. Die Möglichkeit, sich weitere Seiten anzeigen zu lassen, fiel den Testpersonen teilweise gar nicht oder erst sehr spät auf. Diese Möglichkeit, sich die weiteren Seiten anzeigen zu lassen, wurde während der Laboruntersuchungen daher kaum genutzt.

Nach Absenden der ersten Suchanfrage äußerte über die Hälfte der Testpersonen zwar eine Zufriedenheit bezüglich der Anzahl der Ergebnisse, jedoch beurteilten sie diese überwiegend als nicht relevant für ihre gestellte Situationsaufgabe. Aus diesem Grund haben sich auch vier der Testpersonen für eine eigene Frageeinstellung in dieser Situation entschieden. Zum größten Teil betrachteten diese Personen die Ergebnisse als nicht relevant, da sie diese als nicht aktuell genug empfanden. Das betraf überwiegend die Meinungen und Empfehlungen zu technischen Produkten, die älter als sechs Monate waren. Die verbleibenden vier Personen haben ihre Suchanfragen abgeändert, bis ein zufrieden stellendes Ergebnis erzielt wurde oder die Situation von diesen Testpersonen abgebrochen wurde.

2.6 Situationsaufgabe 2: Faktenfragen

Bei dieser Aufgabenstellung (vgl. Anhang 5) waren intuitiv alle Testpersonen der Meinung, dass sich diese Information nicht durch eine Recherche in Lycos IQ ermitteln lässt. Es wurde angegeben, dass unter anderen Umständen als in der Testsituation eine Suchmaschine zur Recherche genutzt werden würde.

Obwohl die Testpersonen keine Ergebnisse erwarteten, haben acht von ihnen eine Recherche auf Lycos iQ durchgeführt; etwa die Hälfte hat „gehofft“, doch noch ein relevantes Ergebnis zu erzielen. Als Begründung wurde genannt, dass man ungern andere Lycos iQ-Nutzer mit der eigenen Frage „belästigen“ möchte. Ebenso wurde seitens der Testpersonen geäußert, dass sie ihre Informationen möglichst sofort haben möchten. Bei der Einstellung einer Frage wurde vermutet, dass es wahrscheinlich länger dauern wird, bis die Frage gelesen und beantwortet wird. Über die Hälfte der Testpersonen erkannten, dass die Antwort von anderen Lycos IQ-Nutzern gegeben wird. Als Antwort wurde von den Testpersonen entweder reiner Text oder Text mit Linkangaben und Quellen erwartet. Zwei Testpersonen haben geäußert, die Antwort in den nächsten Minuten oder Stunden zu erwarten. Weitere zwei gaben an, diese im Laufe der nächsten Tage zu erwarten. Eine Person war der Meinung, dass es länger als eine Woche dauert, bis eine Antwort gegeben wird und eine Person rechnete damit gar keine Antwort auf die gestellte Frage zu erhalten. Zum Prozess der Frageeinstellung siehe Abschnitt 2.8.

Etwa die Hälfte der Testpersonen war mit den erhaltenen Antworten auf ihre eingestellte Frage zufrieden. Die Testpersonen waren, was die Richtigkeit der Information angeht, sehr leichtgläubig und vertrauensvoll. Sie gaben an, die Antwort vor weiterer Verwendung nicht nochmals überprüfen zu wollen. Auch der Rang der Autoren weckte Vertrauen in den Testpersonen. So vermutete eine Testperson, dass der Autor mit dem Rang „Doktorand“ auch „im wirklichen Leben“ Doktorand ist und die Frage in sein fachliches Arbeitsgebiet fällt.

Die Mehrfachnennungen einer Antwort von verschiedenen Lycos iQ-Nutzern wurden nicht als störend empfunden. Die Testpersonen sahen die Richtigkeit der Antwort dadurch bestätigt. Eine Person war mit der gegebenen Antwort nicht zufrieden, da diese nicht zutreffend auf die Fragestellung war. Diese reagierte sehr verärgert und ging dem Bedürfnis nach, dem Autor der nicht zutreffenden Antwort einen Kommentar zu schreiben.

Die restlichen Testpersonen konnten sich nicht zu den Antworten äußern, da diese entweder nicht beachtet oder erst nach Beendigung des Labortests eingegangen sind. Insgesamt waren die Testpersonen aber sehr überrascht, wie schnell ihre Fragen beantwortet wurden.

2.7 Situationsaufgabe 3: Bookmarks

Vor der Untersuchung haben vier Testpersonen angegeben, den Begriff Social Bookmarking zu kennen. Ebenfalls vier Testpersonen gaben an, dass ihnen dieser Begriff nicht geläufig ist. Zwei Personen war der Begriff zwar bekannt, die Bedeutung war ihnen jedoch unklar.

Mit Hilfe der qualitativen Laboruntersuchung sollten Erkenntnisse darüber gewonnen werden, wie die Testpersonen mit den Funktionen von Social Bookmarking umgehen und wie zufrieden sie damit sind. Zu diesem Zweck sollten

sich die Testpersonen in die Situation versetzen, nach geeigneten Websites zu einem bestimmten Thema zu suchen. Um die Suche der Testpersonen gut miteinander vergleichen zu können, wurde allen Testpersonen dieselbe Aufgabe gestellt (vgl. Anhang 5).

Die Suche nach geeigneten Websites sollte zum Vergleich sowohl bei Lycos iQ als auch bei Mister Wong durchgeführt werden. Da die einzelnen Untersuchungen sehr unterschiedlich abliefen und die Testpersonen zum Teil sehr lange mit den vorangegangenen Aufgaben beschäftigt waren, wurde diese Aufgabe nicht allen Testpersonen gestellt. Insgesamt führten sieben der zehn Testpersonen bei Lycos eine Suche nach Websites durch. Die anderen drei Testpersonen lösten die Aufgabe durch das Einstellen einer Frage nach geeigneten Websites. Drei Testpersonen wurden aufgefordert, dieselbe Aufgabe mit Hilfe von Mister Wong zu lösen.

Zunächst wurden die Testpersonen gebeten, auf Lycos iQ nach Websites zu suchen. Diese sind sehr unterschiedlich an die Aufgabe herangegangen (vgl. Abb. 2.5). Vier Personen verwendeten das Suchfeld auf der Startseite von Lycos iQ und gaben dort einzelne Begriffe ein. Zwei Testpersonen wählten dabei die Suchoption „Links“ aus. Eine Testperson führte eine Suche in „ganz Lycos iQ“ durch, da ihr die Option „Links“ unter dem Suchfeld nicht aufgefallen ist. Eine Testperson wählte zunächst die Suchoption „Fragen“ aus, da sie nicht auf die Suchoptionen geachtet hatte und die Option „Fragen“ noch aus einer vorangegangenen Suche markiert war.

Eine Testperson hat die Liste „aktuelle Themen“ für die Suche verwendet. Drei Testpersonen haben als Sucheinstieg den „Linkkatalog“ in der Navigationsleiste ausgewählt. Bei der Suche in der Themenliste bzw. im Linkkatalog haben die Testpersonen eine Anzeige von Unterkategorien erwartet, aus der sie Kategorien für ihre Suche nach Websites auswählen können.

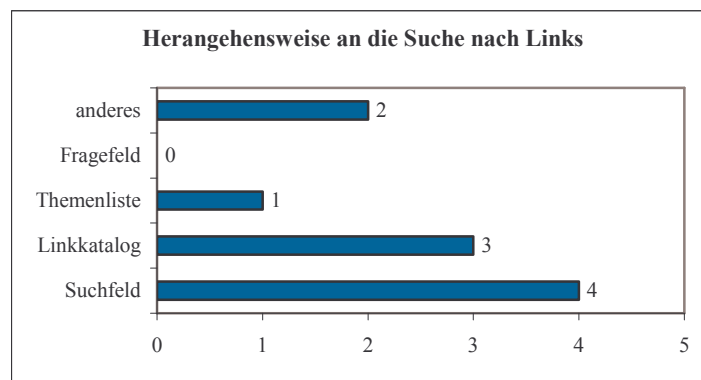


Abb. 2.5 Herangehensweise an die Suche nach Links (n=10)

Von den sieben Testpersonen fanden fünf die angezeigten Treffer relevant. Zwei Personen fanden die angezeigten Treffer nicht relevant bzw. haben keine Treffer erhalten.

Von den sieben Testpersonen haben vier die vergebenen Tags als „Suchwörter“, „Stichwörter“ oder ähnliches erkannt. Die Funktion der Tags für ihre Suche war vier Testpersonen nicht klar. Einer Testperson fiel auf, dass die vergebenen Tags der gefundenen Websites der Trefferliste teilweise nicht zu ihrer Suchanfrage passten.

Die Funktion „kopieren“ fiel keiner der Testpersonen auf. Auf Nachfrage zeigte sich, dass diese Funktion sich aus der Bezeichnung „kopieren“ nicht herleiten lässt, wenn man mit dem Prinzip des Social Bookmarking nicht vertraut ist. Zum Teil wurde vermutet, dass mit dieser Funktion die Trefferliste gespeichert werden kann, um später darauf zurückgreifen zu können.

Eine Testperson vermutete, dass man sich bei Lycos iQ eine persönliche Startseite einrichten kann und dass mit der Funktion „kopieren“ die Treffer auf dieser persönlichen Startseite gespeichert werden können. Nachdem die Testpersonen die angezeigte Trefferliste beurteilt hatten, wurden sie gebeten, einen Link auszuwählen, den sie anklicken würden. Dass die Treffer Links zu externen Websites sind, war sechs der sieben Testpersonen klar. Nur eine Testperson rechnete damit, dass sie beim Klicken auf diesen Treffer auf einer Website von Lycos bleiben würde.

Fünf der sieben Testpersonen wählten einen Treffer mit der Bezeichnung „Anzeige“ aus (vgl. Abb. 2.6). Drei Testpersonen fanden den von Ihnen ausgewählten Link relevant und waren mit diesem Ergebnis zufrieden. Drei Testpersonen fanden den von Ihnen ausgewählten Link für eine weitere Suche hilfreich. Eine Testperson beurteilte den von ihr ausgewählten Link als nicht relevant (vgl. Abb. 2.7).

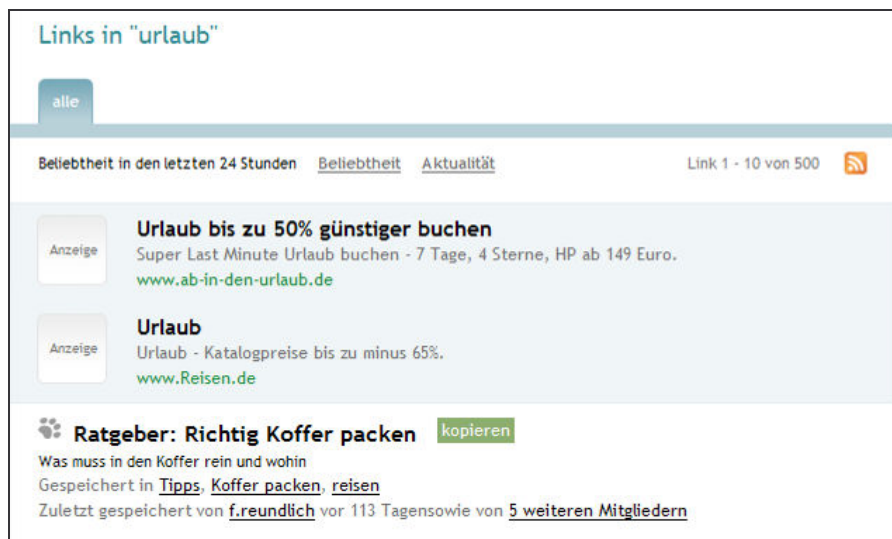


Abb. 2.6 Screenshot der Trefferanzeige bei der Suche nach Links zu „urlaub“

Fünf Testpersonen haben angegeben, dass sie bereit wären, weitere Links aus der Trefferliste auszuwählen. Keine der sieben Testpersonen würde nach dieser ersten Suche weiter bei Lycos iQ nach Websites suchen. Viele gaben an, dass sie die Suche mit einer Suchmaschine weiterführen oder dass sie direkt auf der Website eines ihnen bekannten Reiseveranstalters suchen würden.

Aufgrund der begrenzten Zeit für die Laboruntersuchung und der zum Teil nachlassenden Motivation der Testpersonen wurden nur drei Testpersonen dazu aufgefordert, die Suche nach einer Website bei Mister Wong durchzuführen.

Einer der drei Testpersonen war auf den ersten Blick klar, dass mit Mister Wong nach Websites gesucht werden kann. Die Suchoptionen über dem Suchfeld fielen

dieser Testperson zunächst nicht auf. Die anderen beiden Testpersonen zeigten sich über den Titel „Mister Wong“ irritiert.

Eine Testperson äußerte zunächst die Vermutung, dass bei Mister Wong nach Büchern gesucht werden kann. Dies wurde aus dem Begriff Bookmark geschlossen, da die Testperson den Wortbestandteil „book“ mit Buch übersetzt hat.

Alle drei Testpersonen verwendeten für ihre Suche das Suchfeld auf der Startseite. Nur einer Testperson fiel auf, dass verschiedene Suchoptionen „Websites“, „Favoriten“ und „Tags“ ausgewählt werden können. Die Testperson war dadurch irritiert, da sie die drei Begriffe nicht klar voneinander abgrenzen konnte. Die Testperson führte deshalb intuitiv verschiedene Suchanfragen durch, wobei nur die Suchoption verändert wurde. Warum einige dieser Suchanfragen Ergebnisse lieferten und andere nicht, konnte sich die Testperson nicht erklären. Sie war jedoch damit zufrieden, dass sie schließlich durch Zufall eine Suchoption gefunden hatte, die Treffer erbrachte.

Allen drei Testpersonen war bewusst, dass die angezeigten Treffer Links zu externen Websites waren. Durch die Anzeige der Tags und der Benutzer mit Anzahl konnten die Testpersonen vermuten, dass Mister Wong-Benutzer diese Websites eingestellt und mit Stichwörtern versehen haben.

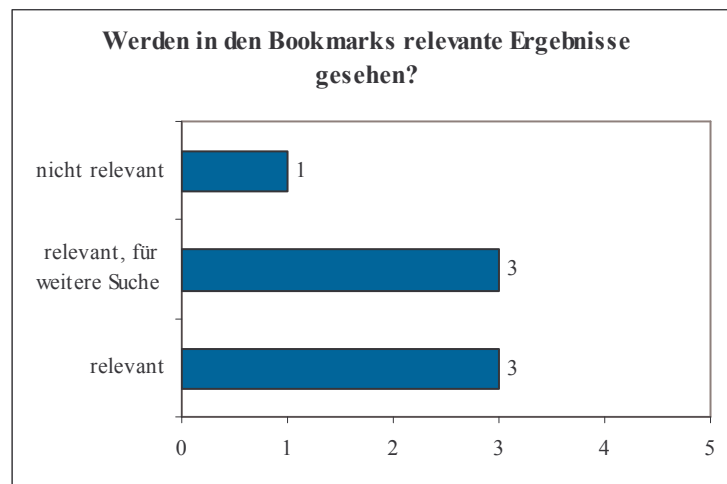


Abb. 2.7 Beurteilung der Relevanz des ausgewählten Links (n=7)

Die Testpersonen wurden aufgefordert, einen Link auszuwählen und diesen nach seiner Relevanz zu beurteilen. Die Sortierung der Treffer erklärte sich den Testpersonen nicht. Eine Testperson vermutete, dass diese nach Relevanz bezogen auf die Suchanfrage sortiert sind. Zum Teil wurden einige der angezeigten Treffer auf den ersten Blick als nicht passend zur vorangegangenen Suche beurteilt. Warum diese in der Trefferliste erschienen, konnte sich die Testperson nicht erklären. Alle drei Testpersonen fanden schließlich relevante Treffer und zeigten sich mit dem Ergebnis zufrieden.

2.8 Einstellen einer eigenen Frage

In diesem Abschnitt wird das Vorgehen der Testpersonen bei der Einstellung einer eigenen Frage näher erläutert. Über die Hälfte der Testpersonen haben diese Funktion über den Link in der linken Menü-Leiste aufgerufen (vgl. Abbildung 2.8). Insgesamt wurde diese Funktion siebzehnmal während der gesamten Laboruntersuchung aufgerufen.

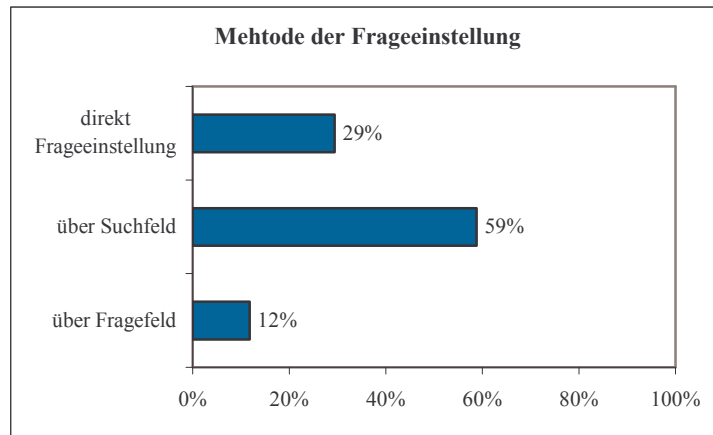


Abb. 2.8 Methode der Frageeinstellung (n=17)

Der Vorgang einer Frageeinstellung unterteilt sich in vier Schritte. Der vierte Schritt der Frageeinstellung variiert, je nachdem ob die Testperson mit einem Account eingeloggt ist. Bei drei Testpersonen war ein Einloggen zum Zeitpunkt der Testdurchführung nicht möglich.

Auffällig war, dass die Hinweistexte zu den einzelnen Schritten größtenteils erstmal nicht beachtet wurden, sondern erst bei der Anzeige einer Fehlermeldung oder bei auftauchenden Unsicherheiten. Den Testpersonen war anfangs teilweise nicht klar, dass sie bei der Einstellung einer Frage nicht sofort eine Antwort erhalten. Die Laboruntersuchung hat auch gezeigt, dass eine gewisse Scheu besteht, eine eigene Frage einzustellen. Somit wurde in elf Fällen vor der Einstellung der Frage erst einmal selbst recherchiert, auch dann, wenn die Testpersonen angaben, keine Treffer zu dem Thema zu finden. Als Grund hierfür wurde genannt, dass man keine Frage stellen möchte, die schon vorhanden ist. Drei der Testpersonen waren der Meinung, sie müssten die Community ansprechen, um eine Antwort auf ihre Frage zu erhalten. Da sowohl Meinungs- als auch Faktenfragen eingestellt wurden, herrschte diese Meinung nur bei den Faktenfragen vor. Im ersten Schritt (vgl. Abb. 2.9) sollte eine Zusammenfassung der Frage eingegeben werden, die dann im zweiten Schritt näher ausgeführt wird. In zwei Fällen aller Frageeinstellungen wurden hier die Stichwörter aus der vorangegangenen Recherche übernommen.

Während der Laboruntersuchung hat die Hälfte der Testpersonen hier ihre Frage ausführlich eingegeben.



***Deine Frage:**

(noch 255 Zeichen verfügbar)

Schritt 1

Gib die Zusammenfassung deiner Frage ein, zum Beispiel "Wo kann ich günstig ein Auto kaufen?". Diese wird später der Titel der Frage sein.

Abb. 2.9 Schritt 1 der Frageeinstellung: Frageneingabe

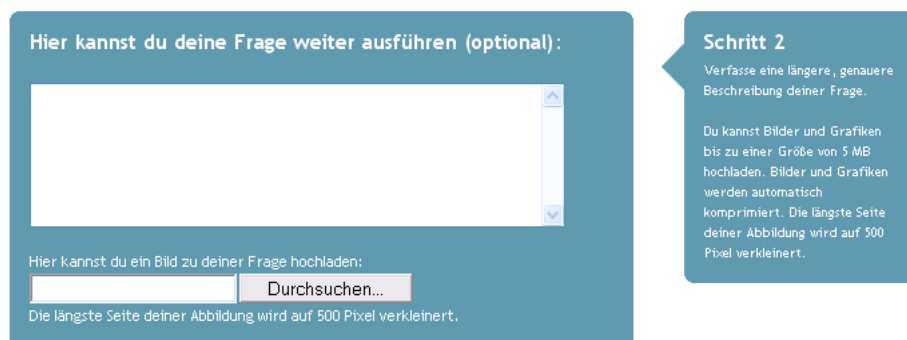
Die Anzeige von schon vorhandenen Fragen wurde in acht Fällen aller Frageeinstellungen gesehen. Vereinzelt wurde auch eine der vorgeschlagenen Fragen ausgewählt und angeschaut. Die Frage enthielt aber nicht die benötigte Information, somit sind die Testpersonen wieder zu ihrer eigenen Frageeinstellung zurückgekehrt.

Auch die Auswahl der angezeigten vorhandenen Fragen war für die Testpersonen nicht ersichtlich. Einige Testpersonen äußerten die Vermutung, dass es bisher nur diese drei Fragen zu der Fragethematik auf Lycos iQ gibt. Diese Personen hätten erwartet, dass schon alle vorhandenen Fragen zu dem Thema angezeigt werden.

Im zweiten Schritt der Frageeinstellung (vgl. Abb. 2.10) kann die Frage aus Schritt 1 weiter ausgeführt werden. Dieses wurde auch von acht Testpersonen getan, zum Teil auch von denen, die ihre Frage im Schritt 1 schon ausführlich gestellt hatten. Die verbleibenden Testpersonen erkannten, dass diese Angabe optional ist und entschieden sich gegen eine weitere Ausführung ihrer Frage.

Die Option ein Bild hoch zu laden war zum größten Teil verständlich. In zwei Fällen wurde nur der Button mit der Beschriftung „Durchsuchen“ angeklickt, ohne die weiteren Erläuterungen zu beachten. Bei diesen Testpersonen entstand die Vermutung, durch das Anklicken dieses Buttons zu den Suchergebnissen zu gelangen. Nachdem sich aber das Fenster mit den Dateien und Ordnern des Testrechners geöffnet hatte, wurde den Testpersonen bewusst, dass sie eine falsche Vermutung angestellt hatten. Das Fenster wurde geschlossen und die Hinweise um den Button herum gelesen. Danach war auch diesen Personen die Funktion verständlich. Eine Testperson hat hier vermutet, ein Bild von sich selbst hochladen zu können und geäußert, dass sie dieses unter keinen Umständen möchte.

Durchgängig haben die Testpersonen die Möglichkeit, ein Bild zur Frage einstellen zu können, als gut bewertet. Jedoch wurde auch angemerkt, dass es für diese fiktiven Situationen nicht sinnvoll wäre.



Hier kannst du deine Frage weiter ausführen (optional):

Hier kannst du ein Bild zu deiner Frage hochladen:

Durchsuchen...

Die längste Seite deiner Abbildung wird auf 500 Pixel verkleinert.

Schritt 2

Verfasse eine längere, genauere Beschreibung deiner Frage.

Du kannst Bilder und Grafiken bis zu einer Größe von 5 MB hochladen. Bilder und Grafiken werden automatisch komprimiert. Die längste Seite deiner Abbildung wird auf 500 Pixel verkleinert.

Abb. 2.10 Schritt 2 der Frageeinstellung: weitere Ausführung der Frage

Im dritten Schritt (vgl. Abb. 2.11) sollte die gestellte Frage einem oder mehreren Themen zugeordnet werden. Hier hat nur knapp die Hälfte der Testpersonen erkannt, dass es sich um eine Pflichteingabe handelt. Die restlichen Testpersonen haben diesen Schritt erst einmal übergangen, was zu einer Fehlermeldung geführt hat. Sechs Testpersonen waren der Überzeugung, dass das Thema der Frage schon in den Schritten eins und zwei angegeben wurde. Die Restlichen vier konnten den Sinn dieser Eingabe erfassen und einen Bezug zu der Themenliste und den schon gesehenen Tags herstellen.

***Die Themen (z. B.: Auto, kaufen, gebraucht) zu deiner Frage:**

Schritt 3

Anhand der Themen können wir Mitglieder ermitteln, die deine Frage beantworten könnten. Diesen wird dann die Frage vorgeschlagen.

Bitte gib mehrere Themen ein, zum Beispiel "auto, kaufen". Mehrere Themen kannst du durch ein Komma trennen.

Abb. 2.11 Schritt 3 der Frageeinstellung: Themenvergabe

Die Testpersonen haben durchschnittlich 1,82 Themen während der gesamten Laboruntersuchungen vergeben. Die Anzeige der schon in Lycos iQ vorhandenen Themen wurde von der Hälfte der Testpersonen wahrgenommen. Somit wurde sich in sechs Fällen bei der Vergabe der Themen an dieser Anzeige orientiert, indem ein Thema durch Anklicken ausgewählt oder bewusst abgeschrieben wurde. In Einzelfällen tauchte auch die Vermutung auf, dass nur die angezeigten Themen vergeben werden können und keine eigenen. Überwiegend haben die Testpersonen allgemeine Begriffe als Thema vergeben und keine speziellen Begriffe, die das Thema weiter eingrenzen würden. Diese stehen, bis auf ganz geringe Ausnahmen, in Bezug zum Inhalt der Fragestellung.

Nur in einem Fall wurde bewusst ein Thema vergeben, das überhaupt nicht zum Inhalt der Fragestellung gehörte. Hier gab die Testperson an, sich über das Anzeigen der Fehlermeldung aufgrund der fehlenden Themen zu ärgern. Die Person war der festen Überzeugung, das Thema schon in der Fragestellung selbst genannt zu haben, da in Feld eins Stichwörter eingegeben wurden. Daraufhin hat die Testperson absichtlich ein nicht passendes Thema vergeben.

Alle Fälle einer Fehlermeldung ließen sich auf diesen Schritt. Entweder waren keine Themen eingegeben oder das Komma, welches die Themen voneinander abgrenzt, war nicht gesetzt. Während einer der Laboruntersuchungen wurde an dieser Stelle auch eine Fehlermeldung gegeben, obwohl die Testperson die Eingabe korrekt ausgeführt hat.

Die Anzeige des Schrittes vier variierte je nachdem, ob die Testperson mit einem Account eingeloggt war. Die drei nicht eingeloggt Testpersonen erhielten die Anzeige mit der Aufforderung sich einzuloggen oder ihre E-Mail-Adresse zu hinterlassen (vgl. Abb. 2.12). Auf das Hinterlassen der E-Mail-Adresse wurde sehr skeptisch reagiert. Die Testpersonen waren sich nicht sicher, inwieweit die E-Mail-Adressen weitergegeben werden und fürchteten Spam-Mails. Daher wurde von diesen Testpersonen die Frageeinstellung an dieser Stelle abgebrochen.

Bitte einloggen oder E-Mail-Adresse hinterlassen

Wenn du bereits bei Lycos IQ als Mitglied registriert bist, melde dich bitte hier an.

Mit Angabe deiner E-Mailadresse stimmst du unseren AGB zu. Wir benötigen deine E-Mailadresse, um dich über eingegangene Antworten zu benachrichtigen und dir weitere Informationen, die dich als Lycos IQ Nutzer betreffen, zukommen zu lassen. Wir werden an deine E-Mailadresse keine sonstige unerwünschte Werbung schicken.

GO

Neuer Benutzer?
Bitte gib hier deine E-Mail-Adresse ein. Wir senden dir eine Aktivierungs-E-Mail für diese Frage sowie eingehende Antworten zu. Diese Benachrichtigungen kannst du später abbestellen.

Vorschau

Abb. 2.12 Schritt 4 der Frageeinstellung ; E-Mail-Adresse

Die sieben mit einem Testaccount eingeloggt Personen erhielten eine Anzeige mit der Einstellung der Laufzeit ihrer Frage und der Auswahl der zu vergebenden Bonuspunkte (vgl. Abb. 2.13). Diese Optionen werden im Abschnitt 2.9 näher erläutert. Überwiegend wurden hier aber die vorgegebenen Optionen übernommen.

Die Buttons mit den Beschriftungen „Vorschau“ und „Speichern“ waren den Testpersonen nicht ganz klar. An dieser Stelle wurde ein Button mit der Beschriftung „Frage einstellen“ oder „Frage absenden“ erwartet. Es konnte aber erahnt werden, welche Funktionen sich dahinter verbergen. In vier Fällen der Frageeinstellungen haben sich die Testpersonen eine Vorschau anzeigen lassen. Diese haben nach der Vorschauanzeige jedoch keine Änderungen an ihrer Frageeinstellung getätigt und sie so abgespeichert.

Ablaufdatum und Bonuspunkte:

Ablaufdatum 7 Tag(e) Bonuspunkte 50

1 7 0 1,000

Schritt 4
Bitte bestimme das Zeitlimit für die Beantwortung der Frage und die eingesetzten Bonuspunkte.

Vorschau Speichern

Abb. 2.13 Schritt 4 der Frageeinstellung: Ablaufdatum und Bonuspunkte

Während der Laboruntersuchungen wurde aber nicht jede ausgefüllte Frageeinstellung abgespeichert, um auf der Lycos IQ-Website zu erscheinen. Zwölf der 17 Frageeinstellungen wurden abgespeichert und somit online gestellt. Das Abbrechen der Frageeinstellung ist darauf zurückzuführen, dass die Testpersonen nicht eingeloggt waren und auch nicht ihre E-Mail-Adresse angeben wollten.

Generell haben die Testpersonen geäußert, dass sie ihre E-Mail-Adresse nicht für die Durchführung einer Informationssuche preisgeben möchten. Das kann aber auf die Testsituation und die gestellte Situationsaufgabe zurückgeführt werden. Die Testpersonen hatten kein wirkliches Interesse an der Information, die in Erfahrung gebracht werden sollte und waren auch alle davon überzeugt, diese Information mittels einer Suchmaschine selbst ermitteln zu können.

2.9 Besonderheiten

Bei der Durchführung der Laboruntersuchung hat sich die jeweilige Moderatorin an einem festgelegten Ablaufschema (vgl. Anhang 3) orientiert, damit sich die verschiedenen Untersuchungen möglichst gut miteinander vergleichen lassen. Die Testpersonen sollten sich jeweils in drei Situationen hineinversetzen und nach Informationen suchen. Allen Testpersonen wurden ähnliche Aufgaben gestellt, um die einzelnen Herangehensweise gut miteinander vergleichen zu können. Die Testpersonen sind sehr unterschiedlich an die gestellten Aufgaben herangegangen. Einige Besonderheiten und überraschende Ergebnisse der Laboruntersuchung werden im Folgenden dargestellt.

Mehrere Testpersonen hatten Schwierigkeiten, zur Lycos iQ-Startseite zurückzufinden. Das Logo von Lycos iQ wurde zum Teil auch von Testpersonen, die angegeben hatten, viel Erfahrung im Umgang mit Internet zu haben, nicht als Link zur Startseite erkannt. Zusätzlich sorgte für Verwirrung, dass der Link „zur Startseite“ zur Startseite von Lycos führt und nicht zur Startseite von Lycos iQ.

Die Liste „aktuelle Themen“ in der rechten Spalte weckte, nachdem ein erster Blick auf die Website geworfen wurde, die Neugier von drei Testpersonen. Die Funktion der Themenliste wurde allerdings zunächst missverstanden. Die Testpersonen hatten die Vermutung, dass sie - bei Klick auf eines der angezeigten Themen - weitere Unterthemen zu diesem Punkt angezeigt bekommen, um ihre Suche weiter einzuschränken. Wie die aktuellen Themen ausgewählt werden und in welcher Reihenfolge sie in der Spalte angezeigt werden, war den Testpersonen nicht verständlich.

Unklarheiten bestanden auch darin, ob bei Auswahl eines Themas und darauf folgender Eingabe in das Suchfeld nur innerhalb des ausgewählten Themas gesucht wird, oder ob die Themenliste nicht für die Einschränkung der Suche über das Suchfeld geeignet ist.

Das Suchfeld für die Themensuche fiel vier Testpersonen auf, dessen Funktion wurde jedoch missverstanden. Die Testpersonen rechneten damit, dass sie nach einem Thema beziehungsweise einer Kategorie suchen können, die in der Themenliste nicht angezeigt wird, um innerhalb dieser Kategorie zu suchen. Mit der Anzeige der Fragen und Antworten zu diesem Suchbegriff waren die Testpersonen an dieser Stelle nicht zufrieden.

Beim Einstellen der Frage wurden die Testpersonen bei Schritt 4 dazu aufgefordert, eine Laufzeit festzulegen. Den Testpersonen war unklar, wie sich die Laufzeit auf die von ihnen eingestellte Frage auswirkt. Es wurde vermutet, dass die Frage nach Ablauf der Laufzeit nicht mehr beantwortet werden kann bzw. keine Bonuspunkte mehr für die Antwort vergeben werden. Einige Testpersonen stellten sich die Frage, ob die Länge der Laufzeit an den Schwierigkeitsgrad der Frage und den Aufwand für die Recherche der Antwort angepasst werden sollte.

Die Testpersonen äußerten sich unterschiedlich dazu, ob sie eine Laufzeit von sieben Tagen für einen kurzen oder für einen langen Zeitraum für die Beantwortung einer Frage hielten. Eine Testperson äußerte, dass es sich bei dem Internet „um ein schnelles Medium“ handle, und hielt deshalb sieben Tage für einen langen Zeitraum. Andere Testpersonen äußerten dagegen, dass sie, wenn überhaupt, erst spät mit Antworten rechneten. Diese Vermutung lag unter anderem darin begründet, dass die Testpersonen nicht einschätzen konnten, wie groß die Community ist, die die Fragen

beantwortet. Insgesamt rechneten fünf Testpersonen mit den ersten Antworten innerhalb von Minuten oder Stunden. Drei Testpersonen vermuteten, dass sie die ersten Antworten frühestens innerhalb einiger Tage erhalten würden.

Beim Einstellen einer Frage bereitete die Option „Bonuspunkte“ einigen Testpersonen Schwierigkeiten. Unklar war insbesondere, welche Anzahl an Bonuspunkten für eine Frage angemessen ist. Die Vorstellungen der Testpersonen waren dabei sehr unterschiedlich. Eine Testperson hielt 50 Bonuspunkte für ziemlich viel, eine andere hielt mehrere hundert Punkte bei einer schwierigen Frage für angemessen. Ob den Bonuspunkten ein Gegenwert gegenübersteht und welche Auswirkungen deshalb die Vergabe von Bonuspunkten hat, war den Testpersonen unklar. Unklar war auch, ob die Bonuspunkte, die für eine Frage vergeben werden, von den eigenen Bonuspunkten abgezogen werden.

Vereinzelt wurden Zweifel geäußert, ob eine geringe Anzahl verbogener Bonuspunkte dafür sorgt, dass man weniger Antworten erhält. Fünf Testpersonen gingen davon aus, dass die Anzahl der Bonuspunkte die Anzahl der erhaltenen Antworten beeinflussen würde. Sechs Testpersonen wählten die voreingestellten 50 Bonuspunkte.

Durch die Anzeige „Systemnachrichten“ zeigten sich einige Testpersonen zunächst verwirrt. Eine Testperson äußerte, dass der Begriff „Systemnachrichten“ missverständlich sei und sehr technisch klinge. Zwei Testpersonen empfanden es als störend, dass die Systemnachrichten oben auf der Seite angezeigt wurden und erst darunter die Treffer einer ausgeführten Suche. Keine Testperson erkannte, dass die Anzeige „Systemnachrichten“ minimiert werden kann. Die Inhalte der Systemnachrichten und die Verknüpfung mit den eingestellten Fragen und den erhaltenen Antworten waren für die Testpersonen verständlich.

In der vorherigen Befragung gaben sechs Personen an, den Begriff „Tagging“ nicht zu kennen (vgl. Abb. 2.14).

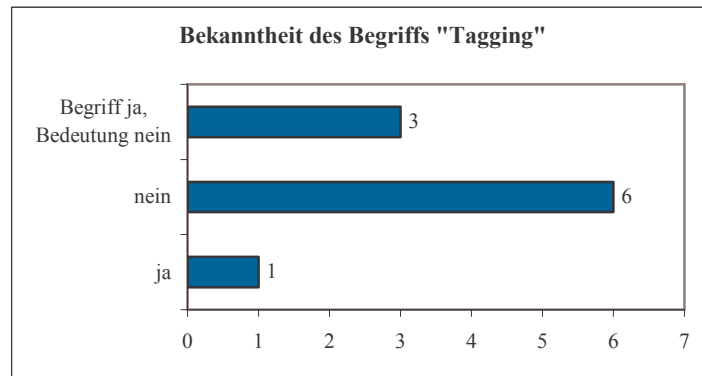


Abb. 2.14 Bekanntheit des Begriffs „Tagging“ (n=10)

Dennoch haben sechs Testpersonen die Tags während der Durchführung der Laboruntersuchung erkannt und auch größtenteils richtig gedeutet. Am häufigsten fielen den Personen die Tags in Situationsaufgabe drei auf. Es war aber nur einer Testperson klar, wer diese Tags vergibt. Einem Teil der Testpersonen wurde erst dann klar, dass die Tags von den Lycos iQ-Nutzern selbst vergeben werden, als sie bei ihrer eigenen Frageeinstellung die Themen vergeben haben.

2.10 Weitere Nutzung von Lycos iQ

Im Anschluss an die Laboruntersuchung wurden die Testpersonen gebeten, einige abschließende Fragen zu beantworten. Die Testpersonen sollten mündlich schildern und begründen, ob und für welche Zwecke sie Lycos iQ zukünftig nutzen würden. Zusätzlich wurden die Testpersonen gebeten, einen kurzen Feedbackbogen (vgl. Anhang 6) auszufüllen und sich zu ihrer Zufriedenheit mit Lycos iQ zu äußern. Die Antwortvorgaben wurden skaliert, um die Auswertung zu erleichtern. Die Ziffern von 3 - 0 stehen für positive Bewertungen wie zum Beispiel „sehr leicht“, „sehr gut“ bzw. „auf jeden Fall“. Die Antwortoption 0 steht für die Mitte, also für „weder – noch“ bzw. „keine Angabe“. Die negativen Bewertungen bewegen sich von -1 (teilweise schlecht) bis -3 (sehr schlecht) (vgl. Anhang 7). Die wichtigsten Ergebnisse werden im Folgenden kurz dargestellt.

Die Testpersonen wurden zunächst gefragt, wie leicht oder schwer es ihnen gefallen ist, auf der Website von Lycos iQ den Bereich zu finden, in dem sie die gewünschte Information erhalten können. Abbildung 2.15 zeigt, dass die Testpersonen überwiegend unzufrieden waren, da sie Schwierigkeiten hatten sich zurechtzufinden.

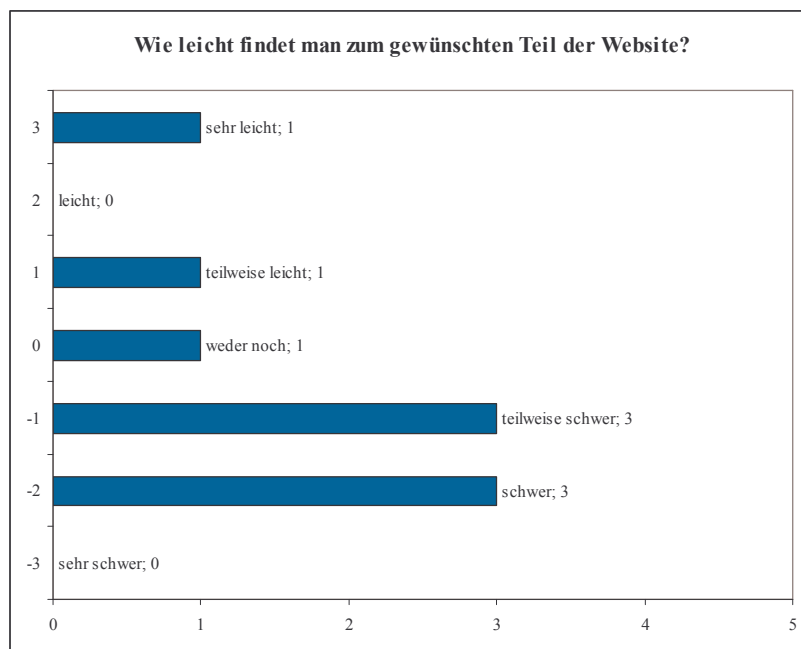


Abb. 2.15 Schwierigkeitsgrad zum gewünschten Teil der Website zu finden (n=9)

Abbildung 2.16 zeigt die Zufriedenheit der Testpersonen mit der Qualität der gefundenen Informationen. Wie diese Grafik zeigt, ist die Mehrheit der Testpersonen mit der Qualität der Informationen unzufrieden. Ein Drittel bewertet die Qualität der gefundenen Information als sehr schlecht.

Die Testpersonen wurden gefragt, welche Funktionen von Lycos iQ sie auch in Zukunft nutzen würden. Abbildung 2.17 zeigt, dass die Möglichkeit, selbst eine Frage einzustellen, bei den Testpersonen die beliebteste Option ist. Für die Suche nach Links würden nur drei der insgesamt zehn Testpersonen Lycos iQ zukünftig verwenden. Ein Grund, der von den Testpersonen häufig genannt wurde, ist die Zufriedenheit mit den Ergebnissen von Suchmaschinen.

Die Testpersonen wurden weiterhin gefragt, bei welcher Art von Fragen und Antworten sie die Nutzung von Lycos iQ bevorzugen. Dabei wurde zwischen Meinungsfragen und Faktenfragen unterschieden.

Als Meinungsfragen wurden Fragen definiert, auf die es keine eindeutig richtigen bzw. falschen Antworten gibt. Andere Lycos-Nutzer werden nach ihrer Meinung und Empfehlungen zu einem bestimmten Thema gefragt.

Als Faktenfragen wurden Fragen definiert, auf die es nur eine richtige Antwort gibt. Dies sind zum Beispiel Fragen nach Jahreszahlen historischer Ereignisse.

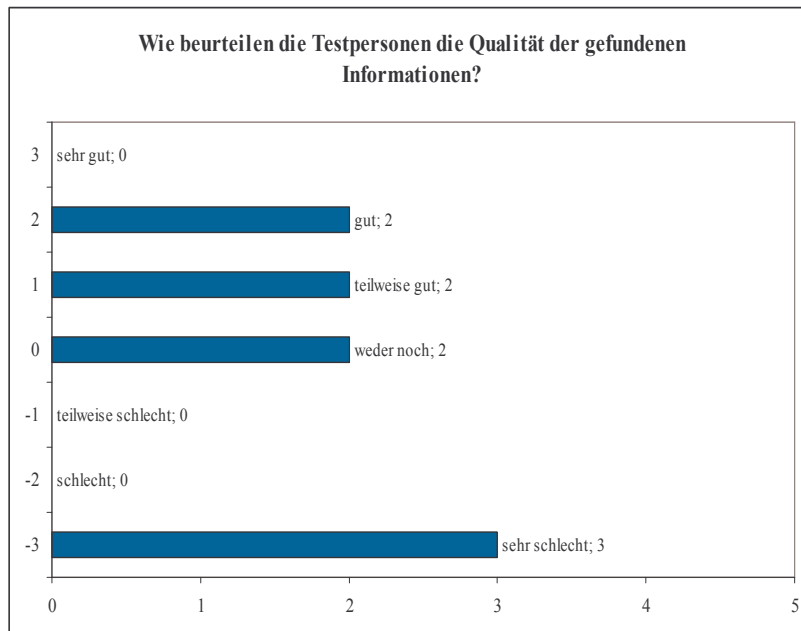


Abb. 2.16 Qualität der gefundenen Informationen (n=9)

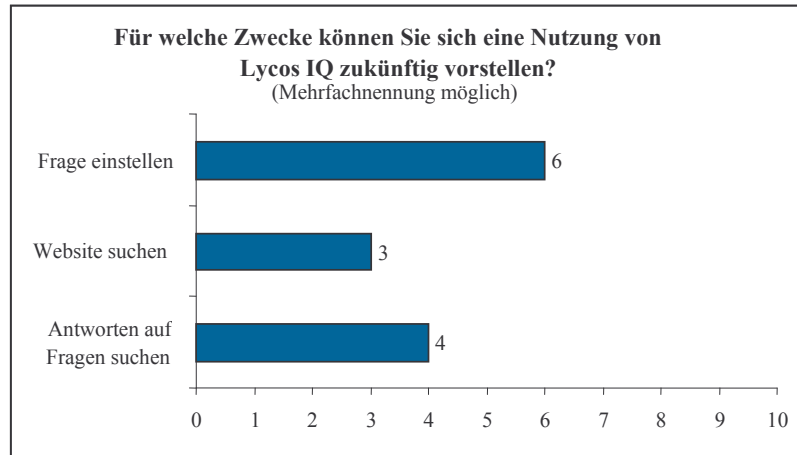


Abb. 2.17 Zukünftige Nutzung von Lycos iQ (n=9)

Die Testpersonen gaben an, dass sie Lycos iQ bevorzugt für die Suche nach Antworten auf Faktenfragen nutzen würden und weniger für die Suche nach Meinungsfragen.

Eine weitere Frage war, warum die Testpersonen Lycos iQ bevorzugt für eine bestimmte Suche nutzen würden. Drei Testpersonen gaben an, dass sie Lycos iQ aus Spaß benutzen würden. Eine Testperson gab an, dass sie Lycos iQ ergänzend zu Suchmaschinen nutzen würde, wenn die Ergebnisse der Suche mit Suchmaschinen nicht zufriedenstellend sind. Eine Testperson gab an, dass sie Lycos iQ aufgrund der Schnelligkeit der erhaltenen Antworten nutzen würde.

Insgesamt waren die Testpersonen mit der Website Lycos iQ überwiegend zufrieden (vgl. Abb. 2.18). Nur eine Testperson bewertete Lycos iQ insgesamt als negativ.

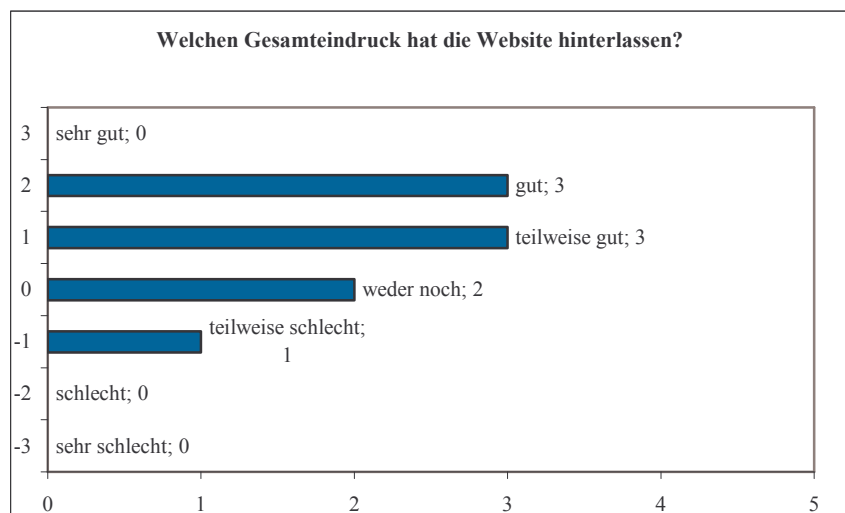


Abb. 2.18 Gesamteindruck von Lycos iQ (n=9)

Insgesamt lässt sich sagen, dass die Testpersonen Schwierigkeiten hatten, auf der Website Lycos iQ den Bereich zu finden, in dem sie die gewünschte Information erhalten. Weiterhin konnte festgestellt werden, dass die Testpersonen mit der Qualität der von ihnen gefunden Informationen überwiegend unzufrieden sind. Trotz dieser negativen Aussagen hat Lycos iQ insgesamt bei zwei Drittel der Testpersonen einen positiven Eindruck hinterlassen.

Die Hälfte der Testpersonen gab an, dass sie Lycos iQ auch zukünftig nutzen würde, wobei die Möglichkeit eine eigene Frage einzustellen, am beliebtesten ist.

3. FAZIT

Ziel der Laboruntersuchung war es, Aufschluss darüber zu erhalten, wie Suchergebnisse von Frage-Antwort- und Social Bookmarking-Diensten in algorithmische Suchmaschinen eingebunden werden können. Im Folgenden werden die inhaltlichen Vorschläge dargestellt und anhand der Laboruntersuchungsergebnisse begründet. Fragen der technischen Umsetzbarkeit sowie rechtliche und eventuell vertragliche Bestimmungen wurden bewusst nicht berücksichtigt.

Die Selektionsmöglichkeit unterhalb der Suchfelder auf der Website Lycos iQ (vgl. Abb. 2.19) wurde von den Testpersonen zu 59 Prozent bewusst ausgewählt. Überwiegend haben die Testpersonen bei Lycos iQ die gesamte Website durchsucht.

Hier wird angeregt, keine Selektionsmöglichkeiten anzubieten. Gegebenenfalls können diese in eine erweiterte Suche ausgelagert werden. Die Laboruntersuchung hat gezeigt, dass die Testpersonen überwiegend die Möglichkeit gewählt haben, die gesamte Website durchsuchen zu lassen, um eine möglichst große Treffermenge zu erhalten. Bei einigen Testpersonen führten die Bezeichnungen der Selektionsmöglichkeiten zu Verwirrung. Beispielsweise konnten sich die Testpersonen nicht erklären, welche Treffer sie bei einer Suche mit der Option „Mitglieder“ erhalten würden. Auch bei Mister Wong war den Testpersonen der Unterschied zwischen den einzelnen Selektionsmöglichkeiten nicht klar (vgl. Abb. 2.20). Eine weitere Schwierigkeit bestand darin, dass sich die Testpersonen zum Teil der voreingestellten Option nicht bewusst waren und aufgrund dessen nur einen Bereich durchsucht haben.



Abb. 2.19 Suchfeld mit Selektionsoptionen bei Lycos iQ

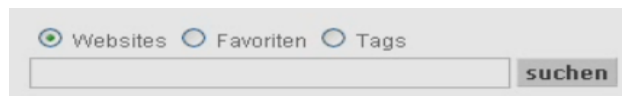


Abb. 2.20 Suchfeld mit Selektionsoptionen bei Mister Wong

Die meisten Testpersonen zeigten bei ihrer Recherche ein Vorgehen, das sie von dem Umgang mit Google her gewöhnt sind. Sie suchten fast ausschließlich mit der Eingabe von einzelnen Suchbegriffen (im Durchschnitt 2,16). Vollständige Sätze oder Phrasen wurden nur in Einzelfällen eingegeben.

Die Möglichkeit Suchbegriffe mit Suchoperatoren („+“) zu verknüpfen wurde nur von einer Minderheit der Testpersonen verwendet. Diese Operatoren sind den Testpersonen aus anderen Suchmaschinen bekannt. Diese haben nicht hinterfragt, ob eine Verwendung von Suchoperatoren bei den getesteten Anbietern überhaupt möglich ist, sondern diese Annahme wurde von den Testpersonen vorausgesetzt. Anhand der durchgeführten Laboruntersuchung kann keine Aussage darüber getroffen werden, ob Social Software-Anwendungen eine Verknüpfung der Suchbegriffe mit Operatoren ermöglichen sollten.

Nach dem Abschicken einer Suchanfrage bei Lycos iQ wurde als Trefferanzeige eine nach Relevanz sortierte Linkliste erwartet, wie es von Suchmaschinen her bekannt ist. Bei der Anzeige der Treffer ist den Testpersonen aufgefallen, dass die Treffer nach den Bereichen Fragen und Antworten, Links und Mitglieder sortiert sind. Nach Absenden der ersten Suchanfrage äußerte über die Hälfte der Testpersonen zwar eine Zufriedenheit bezüglich der Anzahl der Ergebnisse, jedoch beurteilten sie diese als nicht relevant für ihre gestellte Situationsaufgabe. Diese Einschätzung wurde unter anderem mit einem geringen Bezug zur Fragestellung sowie mit mangelnder Aktualität begründet. Die Laboruntersuchung sowie die anschließende Befragung zur Website hat ergeben, dass der Frage-Antwort-Dienst überwiegend für Faktenfragen genutzt werden würde, da die in der Laboruntersuchung erhaltenen Antworten im Vergleich zu Meinungs- und Empfehlungsfragen als relevanter beurteilt wurden.

Obwohl die Mehrheit in der vorangegangenen Befragung angegeben hat, den Begriff „Tagging“ nicht zu kennen, wurden Tags von über der Hälfte der Testpersonen erkannt und ihre Funktion zumindest teilweise richtig gedeutet.

Bei der Vergabe der Tags war auffällig, dass selbst zu speziellen Themen überwiegend allgemeine anstelle von speziellen Begriffen als Tag vergeben wurden. Hier wäre eine Verknüpfung der speziellen mit den allgemeinen Begriffen sinnvoll, da ansonsten bei der Suche nach speziellen Begriffen nur sehr wenig Treffer erzielt würden. Des Weiteren sollten verschiedene Schreibweisen von bedeutungsgleichen Tags automatisch miteinander abgeglichen und gegebenenfalls verknüpft oder angepasst werden, damit relevante Ergebnisse erzielt werden können (vgl. Empfehlungen zur Optimierung des Tagging).

Mehrere Testpersonen beurteilten die Anzeige von Tag-Vorschlägen positiv und vergaben diese bei der Einstellung einer Frage. Tag-Vorschläge sind sinnvoll, wobei diese möglichst speziell sein sollten, um den Inhalt der Frage so genau wie möglich wieder zu geben.

Da während der Laboruntersuchung häufig Tags vergeben wurden, die sich nur wenig oder auch gar nicht auf den Inhalt der Frage bezogen, ist eine Einbindung allein auf Basis der zu einer Frage vergebenen Tags in algorithmische Suchmaschinen fraglich.

Alle zehn Testpersonen haben auf der Website Lycos iQ nach Bookmarks gesucht, davon haben drei dieselbe Suche auch auf der Website Mister Wong durchgeführt. Der Vergleich hat gezeigt, dass die Testpersonen bei keinem der beiden Anbieter einen Unterschied zwischen den Treffern einer Suchmaschine und den Treffern eines Social Bookmarking-Dienstes erkennen konnten.

Insgesamt äußerten die Testpersonen während der Untersuchung, dass sie bei der Suche nach Bookmarks relevante Ergebnisse gefunden haben, mit denen sie zufrieden waren. Beim anschließenden schriftlichen Feedback äußerten jedoch nur drei

Testpersonen, dass sie Lycos iQ zukünftig für die Suche nach Websites nutzen würden. Der hierfür häufig genannte Grund war die große Zufriedenheit der Testpersonen mit den Ergebnissen von Suchmaschinen.

Die Einbindung von Ergebnissen aus Social Software-Anwendungen in algorithmische Suchmaschinen sollte davon abhängig gemacht werden, ob die enthaltene Information schon ausreichend von Suchmaschinen aufgefunden werden kann. Ist dies nicht der Fall, sollte über eine Einbindung nachgedacht werden.

Verwendete Literatur

1. Schulz, Ursula/Spree, Ulrike (Hrsg.): Usability engineering für e-learning: Usability-Tests. Onlinedokument: <http://munin.bui.haw-hamburg.de/amoll/use/methoden/usability-tests.html#3>, [Abruf am 11.01.2008]
2. Machill, Marcel/Welp, Carsten (Hrsg.)(2003): Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen. Bertelsmann-Stiftung: Gütersloh

ANHANG

Anhang 1 – Fragebogen: Hintergrundwissen der Testperson

Datum: _____ Uhrzeit: _____ Test Nr.: _____ User Nr.: _____

Hintergrundbefragung

Danke, dass Sie sich als Freiwilliger für unseren Test zur Verfügung stellen.
Bitte beantworten Sie die folgenden Fragen:

Angaben zur Person

Geschlecht:

- a) männlich
- b) weiblich

Alter: _____

Beruf: _____

Umgang mit dem Internet

1. Wieviele Stunden pro Woche benutzen Sie das Internet? _____ Stunden
2. Von wo aus surfen Sie am meisten?
 - a) Büro
 - b) Zuhause
 - c) Uni
 - d) Sonstiges _____

Spezielle Fragen zum Thema dieses Tests

1. Welche Suchmaschinen nutzen Sie?
 - a) AltaVista
 - b) AOL-Suche
 - c) Google
 - d) Lycos
 - e) Search Live / MSN
 - f) T-Online
 - g) Web.de
 - h) Yahoo!
 - i) andere: _____
2. Welche Frage- und Antwortdienste sind Ihnen bekannt?
 - a) Lycos iQ
 - b) Wer-weiss-was
 - c) Yahoo! Clever

- d) andere: _____
- e) keine

3. Ist Ihnen der Begriff „Social Bookmarking“ (elektronisches Lesezeichen) bekannt?
(z.B. Mr Wong, del.icio.us)?

- a) Ja
- b) Nein
- c) Der Begriff ist mir bekannt, aber die Bedeutung nicht.

4. Ist Ihnen der Begriff „Tagging“ bekannt?

- a) Ja
- b) Nein
- c) Der Begriff ist mir bekannt, aber die Bedeutung nicht.

Vielen Dank für Ihre Unterstützung!

Anhang 2 – Auswertung des Fragebogens

1

a

	männlich	weiblich
Anzahl	6	4

b

Altersgruppe	unter 20 Jahre	20-29 Jahre	30-39 Jahre	40-49 Jahre	ab 50 Jahre
Anzahl	0	4	4	0	2

c

Beruf	Student/in	kaufm. Angest.	Journalist	Pädagogin
Anzahl	3	4	2	1

2

a

Internetnutzung (St. pro Woche)	0-3 Stunden	4-6 Stunden	7-9 Stunden	10-12 Stunden	13-15 Stunden	16-18 Stunden	19-21 Stunden	> 21 Stunden
Anzahl	3	1	1	1	2	1	0	1

b

Ort der Nutzung (hauptsächlich)	Büro	zu Hause	Uni	Sonstiges
Anzahl	5	9	0	0

Anhang 2a - Profil der Testpersonen

			Internetnutzung			Bekanntheit des Begriffs			
Alter	Geschlecht	Beruf	Stunde pro Woche	Ort	genutzte Suchmaschinen	gekante Frage- und Antwortdienste	Social Bookmarking	Tagging	
31	männlich	Industriekaufmann	7	zu Hause	Google, MSN / Life Search, Yahoo!	Wer-weiß-was	nein	nein	
31	männlich	Student	13	zu Hause	Google	Wer-weiß-was	ja	Begriff ja, Bedeutung nicht	
23	weiblich	Studentin	20	zu Hause	Google, Web.de Suche	Wer-weiß-was	ja	nein	
63	männlich	Journalist	2,5	Büro, zu Hause	Google, T-Online Suche, Yahoo!	keine	Begriff ja, Bedeutung nicht	nein	
20	weiblich	Studentin	3	zu Hause	Google	Yahoo! Clever	ja	Begriff ja, Bedeutung nicht	
26	weiblich	Sonderpädagogin	35	Büro, zu Hause	Fireball, Google, Yahoo!	keine	nein	nein	
31	männlich	Diplomkaufmann	11	Büro, zu Hause	Google, Web.de Suche	keine	nein	nein	
62	männlich	Journalist	4	Büro, zu Hause	Google, Yahoo!	Lycos IQ	ja	ja	
23	weiblich	Versicherungskauffrau	2	Büro	Google	keine	Begriff ja, Bedeutung nicht	nein	
31	männlich	Industriekaufmann	10	zu Hause	Google, MSN / Life Search	Lycos IQ, Wer-weiß-was, Yahoo! Clever	nein	Begriff ja, Bedeutung nicht	

Anhang 3 – Ablauf für die aktive Laboruntersuchung

Vorbereitung

Seiten aufrufen und in Account einloggen

Lycos - Benutzername: hawstudi1

für jeden Test muss ein eigener Account angelegt werden (hawstudi1- hawstudi10)

Mr. Wong - Benutzername: hawstudi

der Account kann bestehen bleiben, vor jedem neuen Test müssen ggf. die gesetzten Bookmarks gelöscht werden

Karten mit den jeweiligen Situationsbeschreibungen für die Testperson vorbereiten

Erster Eindruck der Website

- Was gefällt Ihnen an der Website?
- Erster Eindruck der Website
- Farbgestaltung
- Seitenstruktur
- Übersichtlichkeit
- Auffallende Elemente
- Sind die sichtbaren Funktionen verständlich?
- Was wird hinter den Funktionen vermutet?
- Hilfstextnavigation
- Was reizt Sie besonders an der Website? / Worauf würden Sie zuerst klicken?
- Warum reizt es Sie / spricht es Sie an?
- Was erwarten Sie, was passieren wird?
- Was reizt Sie weniger an der Website?
- Was spricht Sie überhaupt nicht an, was finden Sie ausladend?
- Warum?

Aufgabe Frage und Antwortdienste 1 „Erfahrungs-, Meinungsfrage“

(hier wird vermutet, dass die Testperson eine der vorgeschlagenen Fragen zuerst auswählt, daher, für alle Testpersonen dieselbe Situationsbeschreibung)

- Haben Sie eine Vorstellung, wie Sie auf dieser Website an die gesuchte Information kommen können?
- Wie würden Sie vorgehen?
- Vorgehen der Testperson erläutern lassen
- Erwartungen der Testperson erläutern lassen
- Die Testperson die beschriebene Vorgehensweise ausprobieren lassen

Fragenstellung - Erwartungen von der Funktion „Neue Frage stellen“

- Vorgehen erläutern lassen
- Wie würden Sie die Frage formulieren?
- Formulierung erläutern lassen
- Frage eingeben lassen
- Bei der Anzeige der bereits ähnlich gestellten Fragen in Lycos:
 - Was vermuten Sie was das für Fragen sind?
 - Wer hat die angezeigten Fragen formuliert?

- Würden Sie Ihre Formulierung beibehalten oder eine der aufgezeigten Fragen auswählen?
- Warum?
- Bevor die Frage abgeschickt wird:
 - Was vermuten Sie, was passiert?
 - Wer vermuten Sie, beantwortet die Frage?
- Frage abschicken lassen
- Reaktion auf Ergebnis
- Was sehen Sie hier?
- Was interpretiert die Testperson in das Gesehene?
- Welche Elemente fallen der Testperson auf
 - z.B. Bewertungen, Kommentare
 - Statusanzeigen, Datumsanzeigen
- Was vermuten Sie, wer die Antworten gegeben hat?
- Was meinen Sie, warum auf diese Frage geantwortet wurde?
- Wie bewerten Sie die gegebenen Antworten?
- Was meinen Sie, wer die Frage geschlossen hat bzw. warum diese geschlossen wurde?
- In welche Themen-Kategorie würden Sie die Frage einteilen? (Sie dürfen auch mehrere Kategorien nennen)
- Würden Sie nach diesen Antworten Ihre Digitalkamera kaufen gehen?
- Warum / Warum nicht?
- Sind Ihnen die Antworten aktuell genug?
- Beurteilung der Antworten
 - passen zur Frage
 - zufriedenstellend
 - vertrauenswürdig
 - Welche ist Ihrer Meinung nach die beste Antwort?
 - Warum?
 - Welche anderen Antworten hätten Sie erwartet?
 - Welche anderen Antworten hätten Sie sich gewünscht?
 - Hätten Sie sich eine konkretere Antwort gewünscht?
 - Ist für Sie sichtbar, woher die Antworten entnommen wurden?
 - Wie empfinden Sie die Werbung in den Antworten?
- Würden Sie jetzt, nachdem Sie die Antworten gesehen haben, doch Ihre ursprüngliche Frageformulierung noch einmal eingeben oder einer der anderen vorgeschlagenen Fragen auswählen?
- Welche Verbesserungen erwarten Sie von diesen Ergebnissen?
- Wie beurteilen Sie diese Möglichkeit, sich Ratschläge / Meinungen von anderen einzuholen?

Fragestellung – Aufgabe Frage und Antwortdienste 2 „Faktenfrage“

Hier wird vorausgesetzt, dass noch keine passende Frage über genau dieselbe Information besteht → die Testperson muss eine neue Frage stellen. Daher erhält jede

Testperson eine individuelle Situationsbeschreibung, auf die noch keine passende Frage in Lycos iQ gestellt wurde.

Situationsbeschreibung:

- Wie würden Sie vorgehen?
- Vorgehen erläutern lassen
- Erwartungen erläutern lassen
- Ausprobieren lassen
- Fragenstellung
- Vorgehen erläutern lassen
- Wie würden Sie die Frage formulieren?
- Formulierung erläutern lassen
- Frage eingeben lassen
- Beim Aufzeigen der schon ähnlich gestellten Fragen in Lycos:
 - Wird Ihnen wieder eine Frage angezeigt, die sie verwenden können, um an die gesuchte Information zu gelangen?
 - Warum vermuten Sie, wird noch keine passend vorformulierte Frage angezeigt?
- Was erwarten Sie, wenn Sie die Frage abschicken?
- Meinungseinschätzung
- Wie schnell wird eine Antwort vermutet?
- Wie viele Antworten werden auf diese Frage vermutlich gegeben?
- Welche Themen würden Sie für Ihre Frage vergeben?
- Warum würden Sie diese Themen vergeben?
- Welchen Sinn vermuten Sie hinter der Vergabe von Themen?
- Frage abschicken lassen

An dieser Stelle erläutern, dass diese Aufgabe an dieser Stelle beendet ist. Wir können aufgrund des Login mit unserem Account später noch mal prüfen, ob bzw. welche und wie viele Antworten auf die Frage abgegeben wurden und ob diese mit den Erwartungen der Testperson übereinstimmen.

Fragestellung – Bookmarks ; bei Lycos iQ und Mister Wong nach Websites suchen

Für jede Testperson dieselbe Situationsbeschreibung:

Website mit dem entsprechenden Informationsangebot suchen

- Wie können Sie an diese Information gelangen, ohne den Frage- und Antwortdienst zu benutzen?
- „Indirekte“ Hinleitung der Testperson zur Linksuche
- Welche Suchbegriffe würden Sie eingeben?
- Welche Erwartung haben Sie, was passieren wird?
- Eingabe tätigen lassen
- Beurteilung der Ergebnisse
- Können Sie sich vorstellen, wo die Ergebnisse herkommen?
- Ist Ihnen ersichtlich, wer oder was hinter den Links steckt (evtl. Anbieter)?
- Welche Elemente fallen Ihnen auf?

- Was vermuten Sie hinter diesen Elementen?
- „Indirekte“ Hinleitung zu Tags („unterstrichene Wörter“)
- Würden Sie diese Begriffe auch zur Beschreibung des Themas verwenden?
- Warum? / Warum nicht? / Welche Begriffe würden Sie stattdessen bzw. zusätzlich verwenden?
- Wie beurteilen Sie die Treffer?
- Zufrieden? / Unzufrieden?
- Welchen Link würden Sie zuerst anwählen?
- Erläutern Sie Ihre Entscheidung (Warum?)
- Was vermuten Sie, wo sie der ausgewählte Link hinführt?
- Warum vermuten Sie das?
- Link anklicken lassen
- Reaktion
- Beurteilung der Testperson
- Würden Sie Ihre Suchanfragen nach Ansicht dieser Treffer abändern?
- Warum? / Warum nicht?
- Wie würden Sie Ihre Suchanfrage abändern?
- Welche Suchbegriffe würden Sie eingeben?
- Welche anderen Ergebnisse erhoffen Sie sich von der Änderung?
- Würden Sie die vergebenen Tags auch für die Beschreibung der Website verwenden?
- Finden Sie die vergebenen Tags gut ausgewählt oder nicht passend?
- Was meinen Sie, wofür Ihnen die Tags noch nützlich sein könnten?
- Was erwarten Sie, wenn Sie einen bestimmten Tag anklicken?
- Tag anklicken lassen, der nach Meinung der Testperson am besten auf die vorgegebene Situation passt
- Wie beurteilen Sie die jetzige Anzeige?

Praktischer Test ist beendet

Nachfrage:

- Können Sie sich vorstellen diese Site privat oder beruflich zu nutzen?
- Welche Funktionen? / Wofür?
- Warum? / Warum nicht?
- Für was speziell würden Sie Lycos nutzen?
- Wo würden Sie sich bessere Informationen versprechen?
- Welche anderen Ihnen bekannte Websites würden Sie bevorzugen / parallel nutzen?

Anhang 4 – Einverständniserklärung

Hiermit erkläre ich mich einverstanden, dass während der gesamten Laboruntersuchung Ton- und Filmaufnahmen von meiner Person gemacht werden. Die Film- und Tonaufnahmen, sowie persönliche Daten werden ausschließlich von Projektteilnehmern und Mitarbeitern der Firmen Lycos und T-Online genutzt.

Ort, Datum

Name der Testperson
(in Druckbuchstaben)

Unterschrift

Anhang 5 – Testaufgaben

Situation 1 (Meinungsfrage)

1. Sie möchten zu Weihnachten eine Digitalkamera verschenken. Da Sie sich in diesem Bereich nicht auskennen, möchten Sie wissen, welche Hersteller oder welche Produkte von Anderen empfohlen werden.
2. Sie möchten zu Weihnachten eine Digitalkamera verschenken. Da Sie sich auf diesem Gebiet nicht auskennen, möchten Sie wissen, welche Hersteller oder welche Produkte von Anderen empfohlen werden.
3. Sie möchten Ihrer 13-jährigen Nichte eine CD zu Weihnachten schenken. Da Sie aber keine Ahnung haben, was momentan in dieser Altersgruppe angesagt ist, möchten Sie sich Empfehlungen oder Vorschläge von Anderen einholen.
4. Sie möchten gerne mit Ihrem 8-jährigen Neffen einen Kinofilm anschauen. Da Sie aber keine Ahnung haben, welche Kinderfilme zur Zeit laufen und welche altersgerecht sind, möchten Sie sich gerne Empfehlungen von Anderen einholen.
5. Sie möchten zu Weihnachten einen Plasmafernseher verschenken. Da Sie sich auf diesem Gebiet nicht auskennen, möchten Sie wissen, welche Hersteller oder welche Produkte von Anderen empfohlen werden.
6. Sie möchten zu Weihnachten einen Tannenbaum kaufen. Da Sie sich auf diesem Gebiet nicht auskennen, möchten Sie wissen, welche Bäume es auf dem Markt gibt und welche von Anderen empfohlen werden.
7. Sie möchten einen neuen Russfilter für Ihr Diesel-Auto einbauen. Da Sie sich auf dem Markt aber nicht so gut auskennen, möchten Sie wissen, welche Hersteller oder Marken von Anderen empfohlen werden.
8. Sie wollen Silvester in Berlin verbringen. Da Sie sich in der Stadt aber nicht auskennen, möchten Sie gerne von Anderen Veranstaltungstipps und Empfehlungen haben, wo Sie ins neue Jahr feiern können.
9. Sie möchten gerne Ihrem 8-jährigen Cousin zu Weihnachten ein Computerspiel schenken. Da Sie nicht wissen, welche Spiele beliebt und für diese Altersgruppe geeignet sind, möchten Sie wissen, welche Produkte von Anderen empfohlen werden.
10. Sie möchten zu Weihnachten ein Kochbuch mit vegetarischen Rezepten verschenken. Aus diesem Grund möchten Sie wissen, welche Titel von Anderen empfohlen werden.

Situation 2 (Faktenfrage)

1. Im Tatort gibt es einen wortkargen Kommissar, der in Norddeutschland ermittelt. Dieser wird von Axel Milberg gespielt. Sie möchten für ein Quiz in Erfahrung bringen, wie viele Kriminalfälle er in dieser Fernsehserie schon gelöst hat.
2. Sie möchten gerne wissen, wie die weibliche Hauptdarstellerin heißt, die in einem Til Schweiger-Film eine psychiatrische Patientin gespielt hat.
3. In einer amerikanischen Fernsehserie gab es ein intelligentes Tier, das sprechen konnte. Sie möchten gerne wissen wie die Serie heißt und was das für ein Tier war.
4. Sie möchten gerne wissen, wie ein Musical heißt, das vom Leben eines berühmten Schlagerstars handelt und in Hamburg läuft.

5. Sie haben das Zitat „*So einfach wie möglich. Aber nicht einfacher!*“ gehört und möchten gerne wissen, von wem dieses stammt.
6. Sie haben im Internet ein Zitat gefunden, das Ihnen gut gefällt. Leider war auf der Website nicht angegeben, von wem das Zitat stammt. Finden Sie mit Hilfe von Lycos iQ heraus, von wem das folgende Zitat stammt:
„*Das Durchschnittliche gibt der Welt ihren Bestand, das Außergewöhnliche ihren Wert.*“
7. Sie möchten an einem Gewinnspiel teilnehmen und müssen dazu folgende Frage beantworten:
Finden Sie heraus, mit vielen Punkten ein Norddeutscher Fußballverein 1981 Deutscher Meister im Herrenfußball wurde.
8. Sie möchten an einem Gewinnspiel teilnehmen und müssen dazu eine Frage beantworten:
Finden Sie heraus, mit wie vielen Punkten Vorsprung auf den Zweitplatzierten Sandy Shaw 1967 den Grand Prix gewonnen hat.
9. Sie möchten an einem Gewinnspiel teilnehmen und müssen dazu eine Frage beantworten.
Finden Sie heraus, wie viele Einwohner die thailändische Hauptstadt hat.
10. Sie möchten an einem Gewinnspiel teilnehmen und müssen dazu eine Frage beantworten.
Finden Sie heraus, wer laut der aktuellen Ausgabe des Guinness Buchs der Rekorde die größte Sammlung signierter Drumsticks besitzt.

Situation 3 (Suche nach Bookmarks)

Sie möchten nächstes Jahr eine Urlaubsreise nach Dänemark machen. Hierfür suchen Sie nach Websites, die Ihnen Angebote für die Buchung Ihrer Dänemark-Reise bieten.

Anhang 6 – Feedbackbogen Lycos iQ

Datum: _____ Uhrzeit: _____ Test Nr.: _____ User Nr.: _____

Feedback Formular

Bitte bewerten Sie anhand folgender Aspekte Ihre Zufriedenheit mit der Website. Markieren Sie dazu die passende Nummer mit einem Kreis.

- | | | | | | | | | | |
|---|----------------|---|---|---|---|---|---|---|---------------|
| 1. Zum gewünschten Teil der Website finden | sehr leicht | 3 | 2 | 1 | 0 | 1 | 2 | 3 | sehr schwer |
| 2. Qualität der Informationen | sehr gut | 3 | 2 | 1 | 0 | 1 | 2 | 3 | sehr schlecht |
| 3. Die Website kümmert sich um meine Zufriedenheit als Besucher | sehr | 3 | 2 | 1 | 0 | 1 | 2 | 3 | gar nicht |
| 4. Die Angebote auf dieser Website sind relevant relevant für mich. | sehr relevant | 3 | 2 | 1 | 0 | 1 | 2 | 3 | gar nicht |
| 5. Gesamteindruck der Website | sehr gut | 3 | 2 | 1 | 0 | 1 | 2 | 3 | sehr schlecht |
| 6. Würden Sie diese Website nutzen, um Fragen zu stellen? | Auf jeden Fall | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Niemals |
| 7. Würden Sie diese Website nutzen, um nach Antworten auf Fragen zu suchen? | Auf jeden Fall | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Niemals |
| 8. Würden Sie diese Website nutzen, um nach Links zu suchen? | Auf jeden Fall | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Niemals |

Anhang 7 – Auswertung der Feedbackbögen

Nr. zu beurteilen:

1	sehr leicht						sehr schwer	
	Zum gewünschten Teil der Website hinfinden	3	2	1	0	1	2	3
	Anzahl	1	0	1	1	3	3	0

2	sehr gut						sehr schlecht	
	Qualität der Informationen	3	2	1	0	1	2	3
	Anzahl	0	2	2	2	0	0	3

3	sehr						gar nicht	
	Die Website kümmert sich um meine Zufriedenheit als Benutzer	3	2	1	0	1	2	3
	Anzahl	0	0	3	1	1	2	2

4	sehr relevant						gar nicht relevant	
	Die Angebote auf dieser Website sind relevant für mich.	3	2	1	0	1	2	3
	Anzahl	1	2	3	0	1		2

5	sehr gut						sehr schlecht	
	Gesamteindruck der Website	3	2	1	0	1	2	3
	Anzahl	0	3	3	2	1	0	0

6	auf jeden Fall						Niemals	
	Würden Sie diese Website nutzen, um Fragen zu stellen?	3	2	1	0	1	2	3
	Anzahl	2	3	1	1	0	1	1

7	auf jeden Fall						Niemals	
	Würden Sie diese Website nutzen, um nach Antworten auf Fragen zu suchen?	3	2	1	0	1	2	3
	Anzahl	1	3	0	0	0	3	2

8	auf jeden Fall						Niemals	
	Würden Sie diese Website nutzen, um nach Links zu suchen?	3	2	1	0	1	2	3
	Anzahl	0	3	0	1	0	3	2

Anhang 8 – Feedbackbogen Moderation

Datum: _____ Uhrzeit: _____ Test Nr.: _____ User Nr.: _____

Feedback Formular

Ihre Meinung ist uns wichtig. Wie zufrieden waren Sie mit der Durchführung der Laboruntersuchung. Bitte bewerten Sie anhand folgender Aspekte Ihre Zufriedenheit mit der Durchführung dieser Laboruntersuchung.

- | | | | | | | | | | |
|--|----------------|---|---|---|---|---|---|---|-----------------|
| 1. Die Aufgaben waren verständlich und gut durchdacht | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 2. Die Aufgaben waren sowohl von Anfängern als auch für Profis einfach zu lösen. | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 3. Die Moderatorin hat laut und deutlich gesprochen | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 4. Die Anweisungen der Moderatorin waren höflich und bestimmt. | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 5. Das Team machte einen professionellen Eindruck und kannte sich in der Materie aus | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 6. Sie haben sich während des gesamten Testes wohl gefühlt. | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 7. Nach dem Test wurden aufkommende Fragen erläutert | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |
| 8. Der Test hat Ihr Interesse an F&A-Diensten, Social Bookmarking und Tagging geweckt. | Trifft voll zu | 3 | 2 | 1 | 0 | 1 | 2 | 3 | Trifft nicht zu |

9. Der Test war für Sie interessant und informativ	Trifft voll zu	3	2	1	0	1	2	3	Trifft nicht zu
10. Die Durchführung war gut strukturiert	Trifft voll zu	3	2	1	0	1	2	3	Trifft nicht zu

Vergleich der Relevanz von Treffern bei algorithmischen Suchmaschinen, Social Bookmarking-Seiten und Frage-Antwort-Diensten

OLGA GAMMER
olga.gammer@web.de

HEIDI MEISSNER
h_meissner@lycos.de

MAGDALENA PRECKEL
m_preckel@lycos.de

ROBERT OEHLERT
r.oehlert@gmail.com

In diesem Bericht soll die Relevanz der Treffer von algorithmischen Suchmaschinen, Social Bookmarking-Diensten und Frage-Antwort-Diensten untersucht werden. Dazu wird eine Untersuchung mit Probanden durchgeführt, die die Relevanz und einige weitere Kriterien von Treffern bewertet. Die gewonnenen Daten werden statistisch ausgewertet. Sie zeigen, dass alle drei Arten von Suchdiensten eine niedrige Precision aufweisen. Google liefert die relevantesten Treffer, hat jedoch zugleich die meisten kommerziellen Angebote. Social Bookmarking-Dienste liefern besonders viele irrelevante Treffer oder tote Links. Frage-Antwort-Dienste hingegen erzielen eine überraschend gute Relevanz. Es gibt viele Überschneidungen bei den algorithmischen Suchmaschinen und bei den Social Bookmarking-Diensten. Abschließend zeigt die Untersuchung, dass die Suche bei Lycos IQ verbesserungswürdig ist.

1. EINLEITUNG

Um festzustellen, inwiefern die in Lycos iQ gesammelten Informationen zur Verbesserung der Ergebnisrelevanz in die algorithmische Suche eingebunden werden können, wurde beschlossen, zunächst die Relevanz verschiedener Suchdienste untereinander zu vergleichen. Als Grundlage für diese Entscheidung wurde eine empirische Evaluation der Relevanz von Treffern bei verschiedenen Suchdiensten durchgeführt und die daraus resultierenden Daten wurden unter verschiedenen Gesichtspunkten miteinander verglichen. Um die Vielfalt an Suchdiensten zu vergrößern, wurden in die Untersuchung auch Social Bookmarking-Dienste als eine weitere Community-gestützte Alternative zu klassischen Suchmaschinen aufgenommen.

Als problematisch bei dieser Untersuchung stellten sich die seit 40 Jahren in der Informationswissenschaft umstrittene Definition von Relevanz und die Abgrenzung dieser gegen den Begriff Pertinenz [Mizz97] [Stoc07] dar. Da diese Kontroverse

jedoch im Rahmen des Projektes nicht gelöst werden konnte, orientierte sich die Untersuchung an der Definition:

„[R]elevance may be defined as a criterion reflecting the effectiveness of interactive exchange of information between people (or between people and objects potentially conveying information) in communicative relation, all within a context“ [Sara96].

Neben der Precision, welche sich als Quotient aus der Anzahl der gefundenen relevanten Datensätze und der Gesamtzahl der gefundenen Datensätze ergibt [Stoc07], erfasste die Untersuchung verschiedene Aspekte, die helfen können, die Relevanz eines Dokuments zu bestimmen.

Für die Auswertung wurden die Daten der verschiedenen Suchdienste sowohl innerhalb ihrer Kategorie, also nach algorithmischen Suchmaschinen, Frage-Antwort-Diensten oder Social Bookmarking-Diensten, als auch kategorienübergreifend miteinander verglichen, um ein möglichst differenziertes Ergebnis zu erhalten.

Zu Beginn der Untersuchung wurde erwartet, dass gerade die Social Bookmarking-Dienste eine recht hohe Precision erreichen würden, da dort die User selber die Inhalte sammeln. So wie Maaß dachten wir, einen „interessante[n] Ansatz darin zu sehen, die Ergebnisse sozialer Bookmarking[-D]ienste in [...] Suchergebnislisten [von Suchmaschinen] zu integrieren, um die Qualität der Suchergebnisse zu verbessern“ [MaGr07]. Frage-Antwort-Dienste dagegen schätzten wir als keine gute Ergänzung zur algorithmischen Suche ein.

2. METHODIK

Wir haben unseren Relevanztest an drei Suchmaschinen, zwei Social Bookmarking-Diensten und zwei Frage-Antwort-Diensten durchgeführt. Bei den Suchmaschinen entschieden wir uns für Google, Yahoo! und MSN, da sie die drei größten Suchmaschinen mit eigenem Index sind. Bei den Social Bookmarking-Diensten wählten wir Mr. Wong aus, da dies der bekannteste Anbieter dieses Services im deutschsprachigen Raum ist, und del.icio.us, da dieser der bekannteste Anbieter für Social Bookmarking weltweit ist. Aufgrund der engen Zusammenarbeit mit Lycos Europe bezüglich unseres Projektes, war es sinnvoll, deren Frage-Antwort-Dienst Lycos iQ zu in die Analyse mit aufzunehmen. Um vergleichbare Ergebnisse zu erzielen, wählten wir Yahoo! Clever als zweiten Frage-Antwort-Dienst aus. Vom Aufbau und der Grundidee unterscheiden sich diese beiden Dienste kaum.

2.1 Vorüberlegung

Bevor die Untersuchung begonnen hat, galt es, einige Vorüberlegungen zu treffen. Um die Relevanz der Treffer definieren und aufzeigen zu können, musste eine Kriterienliste erstellt werden. Danach folgte die sorgfältige Auswahl der Suchanfragen.

2.1.1 Kriterienliste

Zunächst wurde ein Kriterienkatalog mit fünf Fragen erstellt, die die Relevanz der gelieferten Treffer beurteilen sollten. Dabei war die Frage zu klären, was sich hinter Relevanz verbirgt.

Unter dem Begriff Relevanz lässt sich sowohl die Aktualität der Internetseiten, also der Treffer, als auch deren Verständlichkeit verstehen. Wenn eine Seite unverständlich ist, z. B. aufgrund einer fremden Sprache, wurde sie in der Untersuchung als unverständlich beurteilt. Ein weiteres Kriterium für Relevanz ist der Inhalt der Webseite, d. h. ob ausreichende Informationen enthalten sind, um die Suchanfrage vollständig zu beantworten.

Außerdem ist für den Nutzer die Vertrauenswürdigkeit eines angebotenen Treffers von Interesse, z. B. bei dem Wunsch der Weiterverwendung der Quelle für wissenschaftliche Arbeiten. Relevante weiterführende Links bieten Aufschluss über weitere Informationen. Wie jedoch im Kapitel 2.1.2 beschrieben wird, wurden diese bewusst ausgespart.

Zusammengefasst ergaben sich für jeden Treffer die folgende Bewertungsfragen:

- | | |
|--|------------------------------|
| 1. Ist der Treffer deiner Meinung nach relevant? | Skala 1-6 (Schulnotensystem) |
| 2. Wie schätzt du die Verständlichkeit der Seite ein? | Ja/nein/irrelevant |
| 3. Wie schätzt du die Aktualität der Seite ein? | Ja/nein/irrelevant |
| 4. Ist diese Seite vertraulich? Impressum/Zitierfähigkeit? | Ja/nein/irrelevant |
| 5. Sind weiterführende Links vorhanden? | Ja/nein/irrelevant |

2.1.2 Suchanfragen

Als weiterer Schritt folgt die generelle Auswahl der Suchanfragen. Broder [Lewa05] unterscheidet zwischen drei Arten von Suchanfragen: navigational (navigationsorientiert), informational (informationsorientiert) und transactional (transaktionsorientiert). Mit navigationsorientierten Anfragen sind Webseiten gemeint, die der Suchende bildlich bereits kennt, z. B. wenn er nach populären Unternehmen oder Personen sucht. Informationsorientierte Anfragen bilden die zweite Gruppe, welche über die Informationsvermittlung anhand nur eines Dokumentes hinaus geht. Nach Broder zielen informationsorientierte Anfragen auf jeden Fall auf statische Dokumente, d. h. es ist nach dem Aufruf des Dokuments keine weitere Interaktion auf der Website nötig, um an die gewünschten Informationen zu gelangen [Lewa06]. Hinzu kommen die transaktionsorientierten Anfragen. Transaktion bedeutet hier z. B. ein Produktkauf, ein Datei-Download oder eine Datenbankrecherche.

In der gesamten Untersuchung lag die Konzentration ausschließlich auf den informationsorientierten Suchanfragen.

2.2 Voruntersuchung (Pre-Test)

Vor dem Haupttest galt es, Vorüberlegungen für einen Pre-Test anzustellen. Die daraus resultierenden Ergebnisse sollten in der Hauptuntersuchung berücksichtigt werden.

2.2.1 Selektion der Fragen

Aus einem Logfile der T-Online-Suche mit über 10.000 beliebigen Fragen wurden fünf Fragen nach dem Zufallsprinzip ausgewählt. Ausgesucht wurden nur informationsorientierte Suchanfragen, da eine Relevanzbewertung von transaktionsorientierten Anfragen bei Frage-Antwort-Diensten zu keinem befriedigenden Ergebnis geführt hätte. Da nur bei informationsorientierten Fragen mehrere Dokumente sinnvoll bewertet werden können, ist hier eine zufriedenstellende Relevanzbewertung möglich.

Transaktionsorientierte Fragen dagegen zielen durch ihre kommerzielle Art auf Benutzer ab, die ggf. den Kauf eines Produktes oder die Inanspruchnahme einer Dienstleistung gegebenenfalls beabsichtigen. Es ist unmöglich, mehrere Dokumente sinnvoll zu bewerten, da die abgegebenen Meinungen der Benutzer zu subjektiv sind. Folglich hat der Relevanzbegriff hier eine andere Bedeutung.

2.2.2 Vorgehensweise

Um ein repräsentatives Ergebnis zu erzielen, wurde innerhalb kürzester Zeit jede der fünf Fragen in jedem der sieben Suchdienste gestellt und anonymisiert. Da es sich sowohl um Ein- als auch Mehrwortanfragen handelt, werden die Anfragen vor der Suche in tatsächliche Fragen umgewandelt.

Das Ranking der Treffer sollten die späteren Probanden nicht kennen. Daher erfolgte schon im Pre-Test die Extraktion der Treffer-URLs in ein Excel-Dokument, die sogenannte Anonymisierung der Treffer.

Bei den Frage-Antwort-Diensten wurden auf die vorher gestellte Suchanfrage als „Treffer“ jeweils die ersten drei angezeigten Fragen ausgewählt, eine der dazugehörigen Antworten herauskopiert und der Ursprung unkenntlich gemacht. Dabei wurden nur geschlossene Fragen berücksichtigt. Bei den Antworten wurde die Antwort mit dem „Top-Sternchen“ ausgesucht. Gab es diese nicht, wurde die als „gut beantwortet“ gekennzeichnete Frage genommen. Wenn beide Bewertungen nicht vorhanden waren, blieb nur noch der Griff zum ersten Treffer von oben.

Bewertet werden sollten jeweils die ersten zehn Treffer bei algorithmischen Suchmaschinen und den Social Bookmarking-Diensten. Hinzu kamen drei Fragen pro Frage-Antwort-Dienst inklusive je einer Antwort. Doppelte URLs wurden vermerkt, jedoch ausgeblendet. Die fünf vorher ausgewählten Fragen inklusive anonymisierter Antworten (also Treffer) wurden auf vier Probanden verteilt und nach den Kriterien „Relevante Treffer“, „Verständlichkeit der Seite“, „Aktualität“, „Vertraulichkeit der Seite (Impressum/Zitierfähigkeit)“, „Reichen die Informationen aus?“ und „Sind weiterführende Links vorhanden“ ausgewertet.

2.2.3 Auswertung

Die beim Vortest ausgewerteten Daten waren statistisch nicht aussagekräftig, da sie aus einer viel zu kleinen Stichprobe stammten. Dennoch gaben sie uns einen ersten Eindruck davon, wie unsere Ergebnisse aussehen würden.

Unsere Erwartung, dass Google eine eindeutige Führungsposition einnehmen würde, schien in diesen ersten Ergebnissen bestätigt zu sein. Genauso verhielt es sich

mit unserer Erwartung, dass Frage-Antwort-Dienste eher schlecht abschneiden würden.

2.2.4 Ergebnisse

Erfahrungsgemäß zeichnete sich in den Suchdiensten schnell ab, wenn es keine Treffer zu einer Suchanfrage gab. Die Haupttestvorbereitung begann somit mit der Fragestellung in den Frage-Antwort-Diensten, gefolgt von den Social Bookmarking-Seiten. Gab es zu einer Suchanfrage in einem der ausgewählten Suchdiensten keinen einzigen Treffer, so wurde die Suchanfrage komplett herausgenommen und nicht berücksichtigt. Die Untersuchung war nur dann sinnvoll, wenn in jedem der drei Arten von Suchdiensten Treffer angezeigt wurden, da die Ergebnisse sonst verzerrt gewesen wären.

Wesentliche Erkenntnisse zeichneten sich bei dem Frage-Antwort-Tool Lycos iQ ab. Es erfolgte weder ein Ranking der Antworten noch eine Unterscheidung von Singular- und Pluralbegriffen. Das Fragenranking erschien völlig unklar. Zudem wurde keine Schreibkorrektur oder Phrasensuche angeboten. Vermutlich würde sich ein Ranking der Antworten, z. B. nach „Top-Sternchen“, bei einer zukünftigen Einbettung von Lycos iQ in die Suchmaschinen als tauglich erweisen.

Da Lycos iQ zeitweilig nicht mehr auf der Webseite von T-Online aufgeführt wird und um Komplikationen beim Haupttest zu vermeiden, sollte dieser direkt über die Startseiten der Frage-Antwort-Dienste Lycos iQ und Yahoo! Clever durchgeführt werden. Im Haupttest wurden bei den beiden Frage-Antwort-Diensten die ersten drei Fragen zu der jeweils gestellten Suchanfrage als „Treffer“ ausgewählt, in eine Exceltabelle kopiert und anonymisiert. Da bei den Frage-Antwort-Diensten lediglich die Fragen unsere „Treffer“ waren, blieben die Antworten unberücksichtigt und wurden weggelassen.

Wir hatten uns zu dieser Vorgehensweise aus mehreren Gründen entschlossen. Zum einen werden bei den Frage-Antwort-Diensten nicht die Antworten, sondern die Fragen indexiert. Wenn man eine Suchanfrage stellt, werden die Fragen nach möglichen Übereinstimmungen überprüft und nicht die Antworten, was auch logisch erscheint, denn sonst würde zu viel Informationsballast generiert. Zum anderen werden die Fragen als Trefferliste ausgegeben. Wenn die Frage nicht relevant zur Suchanfrage ist, können die Antworten logischerweise auch nicht relevant sein. Auch präsentiert sich ein Problem bei der Auswahl der Antworten, da es zur Zeit unserer Untersuchung kein Ranking der Antworten gab und mehrere Antworten als „Top Antwort“ gekennzeichnet wurden. Wenn man jedoch alle Antworten aufgenommen hätte, wäre es zu einem Information-Overload gekommen. Des Weiteren wird die Qualität der Antworten bei Lycos iQ von einer anderen Gruppe des Projektes noch einmal ausgiebig beleuchtet (siehe den Beitrag „Untersuchung der Qualität der Antworten bei Lycos iQ und deren Einbindung in die algorithmische Suche“ in diesem Band).

Bei den sieben Suchdiensten waren die unterschiedlichen Ranking-Maßnahmen (Anzeige mit oder ohne Nummerierung) und diverse Zusatzangebote wie beispielsweise kommerzielle Anzeigen, Verbesserungs- oder Tagvorschläge auffällig. Ein zusätzlicher Problempunkt findet sich bei den Social Bookmarking-Seiten wieder. Bei del.icio.us z. B. werden die Links verdeckt angezeigt, d. h. die eigentlichen URLs

sind nicht direkt sichtbar, sondern werden mit einem beliebigen Namen belegt, der im Web für die Nutzer unsichtbar ist und erst bei Anklicken durch das Öffnen der jeweiligen Internetseite im Browserfenster erscheint. Dies führt zu doppelten oder toten Links. Aus Usability-Sicht sind das Anklicken jedes Treffers und das Kopieren der eigentlichen URL in ein neues Browserfenster recht zeitaufwändig und umständlich. Tote Links wurden im Haupttest trotzdem mit aufgeführt und den Probanden gegeben, da die Seiten möglicherweise nur kurzfristig nicht aufrufbar waren.

Ein Problem stellte sich auch bei aufgeführten Unter-URLs, die eigentlich mit Main- oder Index-URLs verwandt sind – den sogenannten „related Links“. Diese Links sind ähnlich strukturiert wie diejenigen auf den Hauptseiten. Die „Unter“-Webseiten beinhalten jedoch meist die eigentlichen Informationen, während die Hauptseiten oft nur Einführungs- bzw. Überblickseiten anbieten und eher zur Navigation als zur eigentlichen Informationsvermittlung dienen. Die Entscheidung, diese Links im Haupttest weiterhin aufzuführen, erschien aus Objektivitäts- und datenmengentechnischen Gründen sinnvoll. Hätten wir diese Links aus unserer Untersuchung entfernt, hätte sich die Auswahl an möglichen Suchanfragen um ein vielfaches minimiert.

2.3 Haupttest (Main-Test)

Um die Relevanzbewertung möglichst objektiv zu gestalten, musste der Relevanztest von unabhängigen Testpersonen durchgeführt werden, die eine hohe Kompetenz im Bereich der Web- und Suchmaschinennutzung besitzen. Alle Probanden kamen demnach aus dem Bereich der Informationsvermittlung. Anhand von zufällig ausgewählten Suchanfragen aus dem Logfile der T-Online-Suche sollten sie die Relevanz von Treffern bei algorithmischen Suchmaschinen, Social Bookmarking-Seiten und Frage-Antwort-Diensten bewerten. Die dadurch gewonnenen Daten werden ausgewertet und interpretiert.

2.3.1 Suchanfragen

Es sollte eine möglichst große Anzahl an Suchanfragen, jedenfalls mindestens 50, ausgewertet werden, denn je höher die Anzahl der Anfragen, desto geringer ist die Fehlerwahrscheinlichkeit bei den Ergebnissen der Evaluation [Grie00]. Von circa 72 Fragen wurden 54 ausgewählt.

Jede Frage wurde in jedem der sieben Suchdienste einmal gestellt (siehe auch 2.2 Voruntersuchung (Pre-Test), 2.2.2 Vorgehensweise) und die jeweiligen Treffer (die Antworten) herausgefiltert. Pro Frage wurden dabei bei den drei algorithmischen Suchmaschinen und den zwei Social Bookmarking-Diensten jeweils die ersten zehn Treffer, bei den zwei Frage-Antwort-Diensten jeweils die ersten drei Treffer betrachtet.

Die Fragen für den Haupttest, die in einem der Suchdienste keinen Treffer erzielten, wurden nicht an die Probanden weitergegeben, sondern nur vermerkt. Eine Einschränkung auf bestimmte Themenbereiche fand bis auf pornographische Themen nicht statt. Die Herkunft der Treffer wurde unkenntlich gemacht, um eine Verzerrung

der Ergebnisse aufgrund von Vorlieben oder Abneigungen bei den Testpersonen zu bestimmten Suchdiensten zu vermeiden [GRB02].

Die Suchanfragen wurden schriftlich ausformuliert, um sicherzustellen, dass die informationsorientierte Seite der Frage tatsächlich im Vordergrund steht. Besonders die Suchanfragen, die nur aus einem Suchbegriff (Einwort-Suchanfragen) bestanden, liessen mehrere Interpretationen zu.

2.3.2 Vorgehensweise und Bewertung der Probanden

Insgesamt beschäftigten sich 20 Juroren mit jeweils ca. 150 Webseiten und neun Fragen aus Frage-Antwort-Diensten, die sich aus Suchanfragen ergeben hatten. Der zeitliche Aufwand pro Suchanfrage wurde auf eine Stunde geschätzt. Jede Suchanfrage und jeder Treffer waren von nur einer Person zu bewerten, um die Eindeutigkeit der Bewertung zu garantieren.

Alle URLs der Treffer der verschiedenen Suchdienste und die Fragen der Frage-Antwort-Dienste wurden mit den entsprechenden Informationen („Von welchem Suchdienst stammt der Treffer?“, „Welchen Rankingplatz hat der Treffer inne?“, „Zu welcher Suchanfrage gehört er?“ und „Handelt es sich hierbei um einen doppelten Treffer?“) in einer Datenbank gespeichert.

Am Testtag wurden die URLs mit einer Identifikationsnummer und den zu bewertenden Kriterien in Excel-Tabellen ausgelesen. Für jede Suchanfrage ergaben sich zwei Tabellen: eine mit den URLs der Treffer und eine mit den kopierten Fragen der Frage-Antwort-Dienste. Dieser Vorgang hatte zum Ziel, den Prozess zu anonymisieren. Um die Treffer noch weiter zu anonymisieren, wurden die URLs alphabetisch geordnet und die Identifikationsnummern ausgeblendet. Die doppelten Treffer und Überschneidungen wurden schon beim Auslesen der Daten aus der Datenbank herausgefiltert, so dass jede Webseite nur einmal bewertet werden musste.

Die Probanden erhielten eine kurze mündliche Einweisung zum Thema Evaluation und ihren Aufgaben als Juroren. Danach bewerteten sie anhand der Bewertungsliste die anonymisierten Treffer. Während der Relevanzbeurteilung sollten die Probanden nicht beeinflusst werden. Die Leiter der Untersuchung standen jedoch für Rückfragen zu technischen Problemen oder für Verständnisfragen zur Verfügung [Grie00].

3. AUSWERTUNG DES HAUPTTESTS

In der Regel wird die Relevanz von Treffern mit dem klassischen Retrievalmaß Precision bewertet. Diese gibt den Anteil der gefundenen relevanten Treffer an der Gesamtzahl der gefundenen Treffer an [StSt00].

Dieses Maß ist verhältnismäßig leicht zu bestimmen, indem alle gefundenen Treffer von den Juroren bewertet werden und anschließend eine Auszählung der relevanten bzw. irrelevanten Treffer erfolgt. Ein Problem taucht aber auf, wenn man anstatt eines zweiwertigen Relevanzurteils (relevant/nicht relevant) eine Skala verwendet, die eine „zusätzliche Qualitätsabstufung möglich macht“ [LeHö07]. Aus diesem Grund wurden in der Untersuchung drei Abstufungen der Precision in „stark“, „mittel“ und „schwach“ unterschieden.

Su, Chen und Dong verwenden bei ihrer Evaluation von Suchmaschinen aus Anwendersicht ebenfalls eine Dreierskala (relevant, teilweise relevant, nicht relevant).

Anschließend rechnen sie sowohl die Precision der relevanten und teilweise relevanten Dokumente zusammen („schwache Precision“) als auch nur der relevanten Treffer („starke Precision“) [StSt00].

3.1 Top Ten Precision

In der Untersuchung der Relevanz von Treffern bei algorithmischen Suchmaschinen, Social Bookmarking-Seiten und Frage-Antwort-Diensten wurden circa 50 Suchanfragen verwendet, für die jeweils die ersten zehn Treffer jedes Suchdienstes (bei Frage-Antwort-Diensten jeweils die ersten drei Treffer) ausgewertet wurden. Die 20 Juroren bewerteten Treffer auf einer Skala von eins bis sechs (angelehnt an das Schulnotensystem: eins = Treffer ist relevant; sechs = Treffer ist nicht relevant).

Für die Auswertung wurde die „starke“, „mittlere“ und „schwache“ Precision berechnet. Allen Treffern, die mit eins bewertet wurden, wurde eine „starke“ Precision zugeschrieben. „Mittlere Precision“ haben diejenigen Treffer, die mit eins und zwei bewertet wurden. „Schwache“ Precision bezeichnet relevante, teilweise relevante und wenig relevante Treffer.

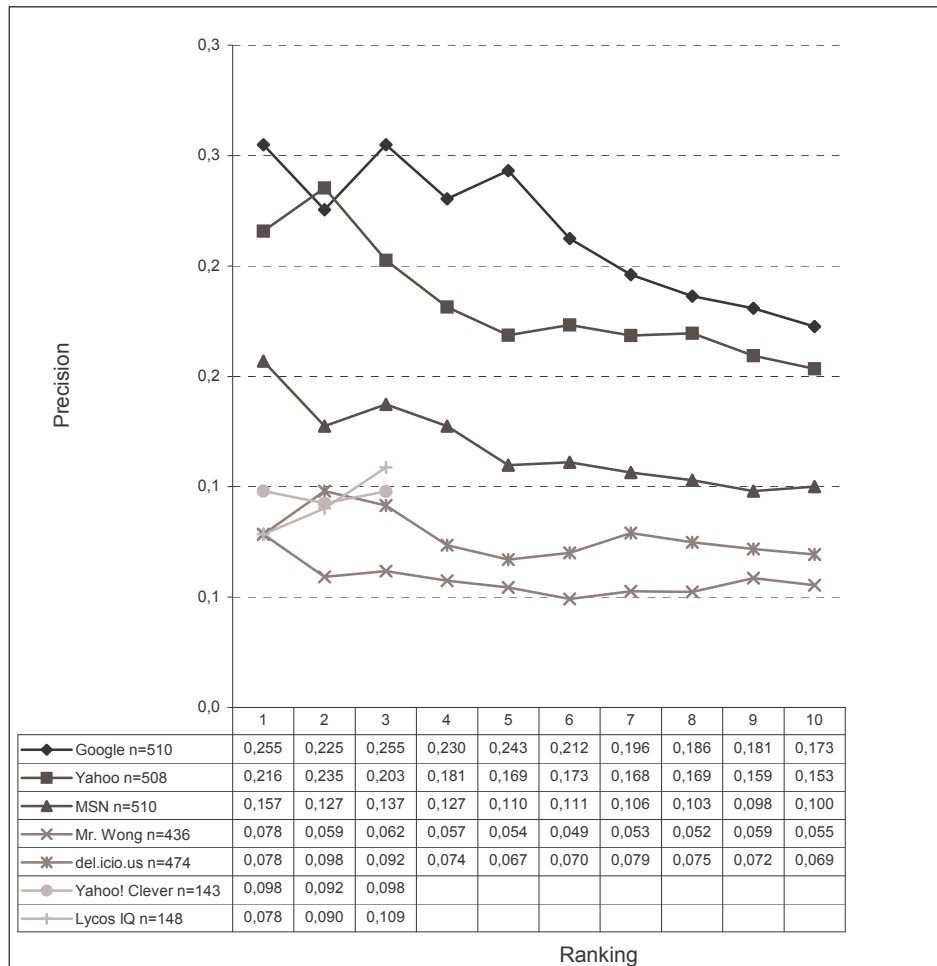


Abb. 3.1 Top Ten Precision („starke“ Precision)

Die Betrachtung der Top Ten Precision („starke“ Precision, Abb. 3.1) zeigt, dass Google fast bei allen Rangplätzen höhere Werte erreichte als die anderen Suchdienste. Eine Ausnahme bildet Rang zwei. Die Kurve überschneidet sich mit der von Yahoo!.

Vergleicht man die Kurven von del.icio.us und Mr. Wong, wird ersichtlich, dass del.icio.us besser abschneidet. Die Kurve von Lycos iQ steigt während der ersten drei Positionen an, während die Kurve von Yahoo!Clever gleich bleibt.

Im Falle der „starken“ Precision sind die Ergebnisse der einzelnen Suchdienste schwer zu vergleichen, weil nur wenige Treffer für die Auswertung vorhanden sind. Es waren lediglich elf Prozent Treffer, die ein Top-Ergebnis lieferten bzw. mit eins bewertet wurden. Besser lassen sich die Ergebnisse der Grafik „mittlere“ Precision auswerten. Insgesamt sind es 25 Prozent der Treffer, die mit eins und zwei bewertet wurden.

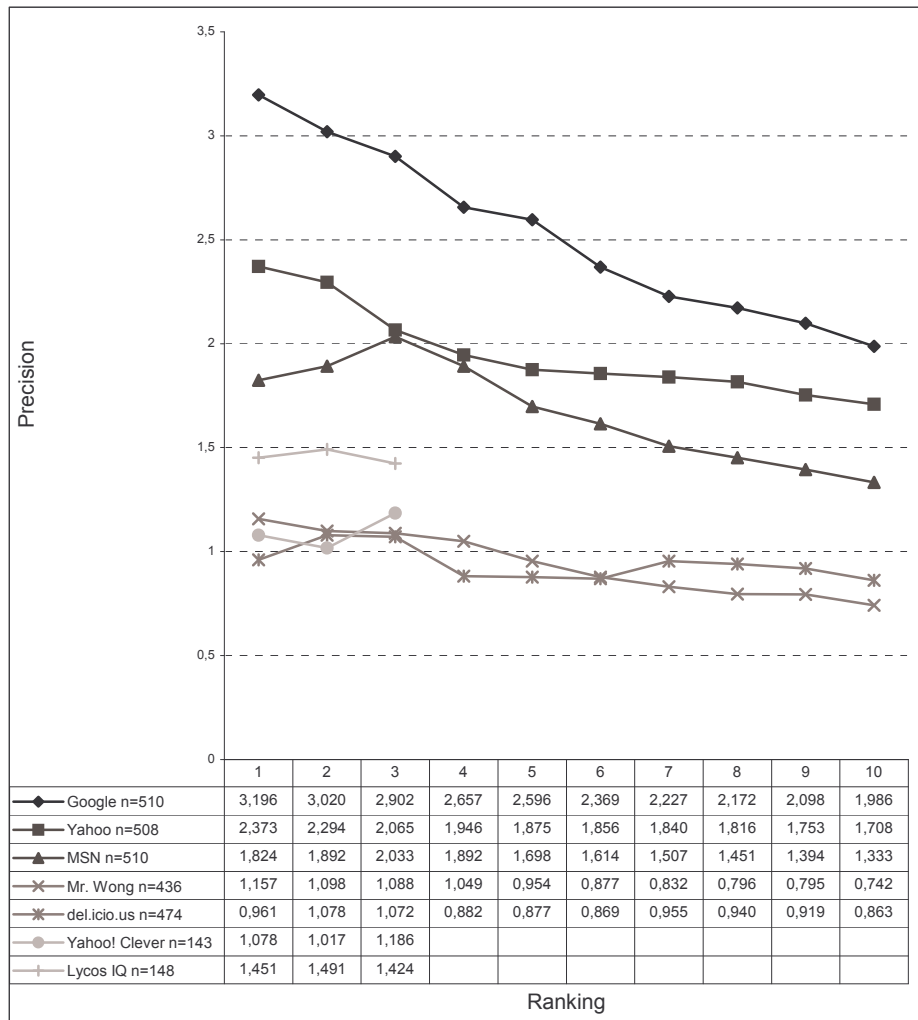


Abb. 3.2 Top Ten Precision („mittlere" Precision)

Für die Auswertung der „mittleren“ und „schwachen“ Precision wurde das Gewicht der einzelnen Treffer berücksichtigt. Die Treffer, die mit eins bewertet wurden, erhielten sechs Punkte, da sie die besten Treffer waren. Die Treffer, die mit zwei bewertet wurden, erhielten fünf Punkte usw.

Der Verlauf der Kurven bei der Messung der mittleren Precision (Abb. 3.2) ist übersichtlicher und eindeutiger dargestellt (vgl. Abb. 3.2). Obwohl die Graphen sich bisweilen überschneiden, zeigt sich klar, welche Suchdienste die relevantesten Treffer lieferten. So erreichte Google bei jedem Rang höhere Precisionwerte als alle anderen Suchdienste. Die Graphen der Social Bookmarking-Diensten überschneiden sich mehrmals, so dass nicht eindeutig ersichtlich wird, welches der beiden Systeme eine

höhere Effektivität erreichte. Bei dem Vergleich der Frage-Antwort-Dienste ließ sich festhalten, dass Lycos IQ besser abschnitt als Yahoo! Clever. Auffällig ist hier der steigende Kurvenverlauf von Yahoo! Clever.

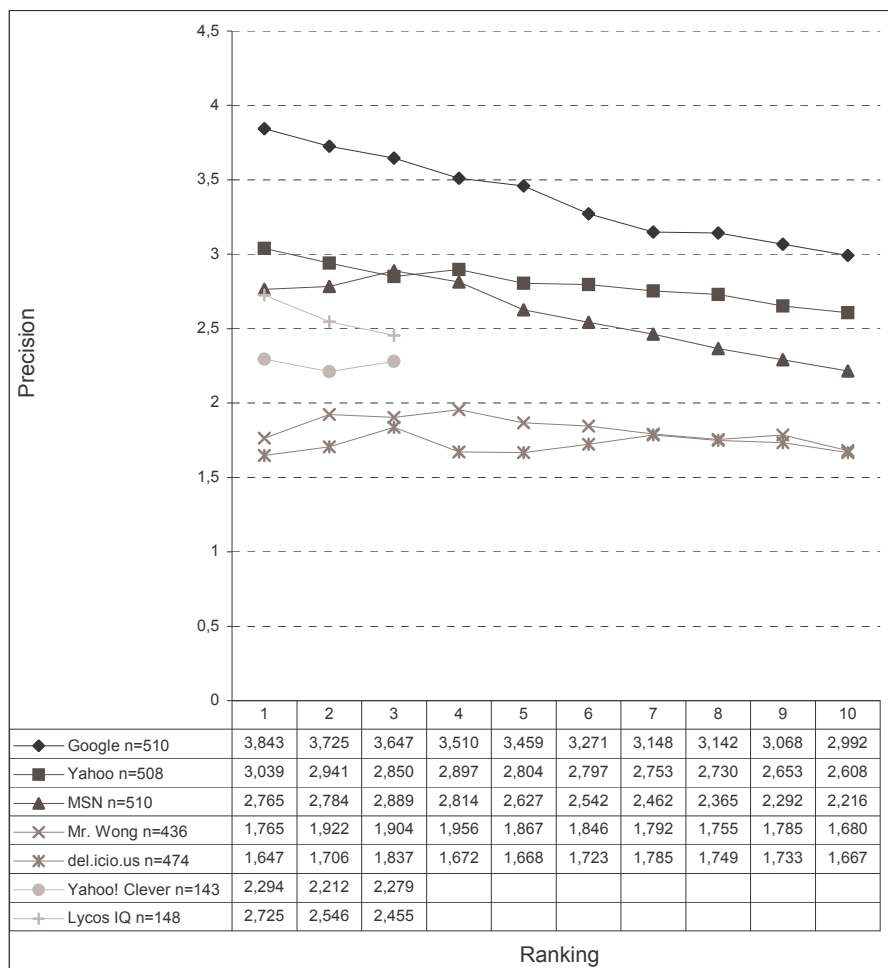


Abb. 3.3 Top Ten Precision („schwache“ Precision)

Mr. Wong erzielte bei der Auswertung der „schwachen“ Precision leicht höhere Werte als del.icio.us, allerdings waren die Unterschiede relativ gering. Die Kurven von Yahoo! und MSN überschneiden sich bei der Position drei, sonst erreichte Yahoo! bessere Precisionwerte als MSN. Bei dem Vergleich der Frage-Antwort-Dienste schneidet Lycos iQ eindeutig besser ab als Yahoo! Clever.

Die Auswertung der „schwachen“ Precision zeigt eine eindeutige Reihenfolge im Hinblick auf die Relevanz der Treffer. Die algorithmischen Suchmaschinen erwiesen sich als die effektivsten Retrievalsysteme, gefolgt von den Frage-Antwort-Diensten. An letzter Stelle stehen die Social Bookmarking-Dienste.

Das Gesamtergebnis über die Precision der ersten zehn Treffer ist in der nachfolgenden Tabelle zu sehen.

Suchdienst:	„starke“ Precision	„mittlere“ Precision	„schwache“ Precision
Google n=510	0,22	0,46	0,75
Yahoo! n=508	0,18	0,36	0,62
MSN n=510	0,12	0,30	0,57
Mr. Wong n=436	0,058	0,176	0,40
del.icio.us n=474	0,077	0,173	0,38
Yahoo! Clever n=143	0,096	0,20	0,50
Lycos iQ n=148	0,092	0,26	0,57

Abb. 3.4 Precision der ersten zehn Treffer

Die Ergebnisse zeigen, dass zwischen den Suchdiensten klare Unterschiede vorliegen. Ungeachtet des Precisionsgrades („stark“, „mittel“ oder „schwach“) belegte Google immer den ersten Platz. Mit circa 46 Prozent der relevanten Treffer bei „mittlerer“ Precision z.B. erzeugte Google ca. 15 Prozent mehr relevante Treffer als MSN.

Beim Vergleich zwischen den algorithmischen Suchmaschinen mit den Social Bookmarking-Diensten und den Frage-Antwort-Diensten lagen die Frage-Antwort-Dienste überraschenderweise an zweiter Stelle, wobei Lycos iQ um sechs Prozent besser abschnitt als Yahoo! Clever. Die letzte Position belegten die Social Bookmarking-Dienste, die im Vergleich zu Google weniger als die Hälfte an relevanten Treffern aufzeigen (ca. 17 Prozent).

Die Gründe, warum die Social-Bookmarking-Dienste so wenige relevante Treffer lieferten, sind:

- viele „tote“ URLs, die automatisch mit der Schulnote sechs bewertet wurden. Dies zeigt, dass nichtfunktionierende URLs aus dem Index der Social Bookmarking-Dienste nicht gelöscht werden. Die Ergebnisse zeigen, dass etwa fünf Prozent der ausgegebenen Treffer „tote“ URLs waren. Deswegen ist es erforderlich, „dass Anwender entsprechende Links in ihre Bookmark-Liste einpflegen“ [MaGr07].
- viele fremdsprachige Seiten (ausgeschlossen sind englischsprachige Seiten), die natürlich zwar relevant sein können, aber aufgrund des Nichtverstehens der Sprache von den Juroren meistens subjektiv als nicht relevant bewertet wurden. Von den 17 fremdsprachigen Seiten, die ausgewertet wurden, stammten 16 aus del.icio.us.

In der folgenden Auswertung wird auf nicht-relevante Treffer und die Gründe für ihre Irrelevanz, sowie auf die Verständlichkeit der Seiten eingegangen. Zur Berechnung des Gesamtergebnisses wurde die Grafik „mittlere“ Precision (Abb. 3.5) herangezogen.

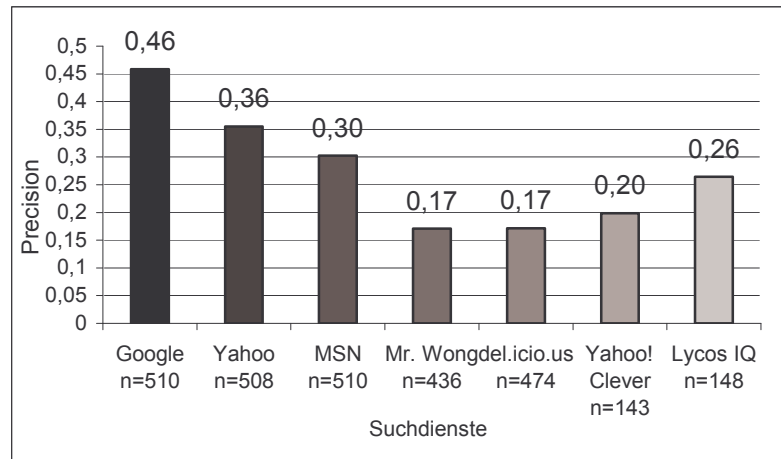


Abb. 3.5 Gesamtergebnis: „mittlere“ Precision

Bei den Ergebnissen der Precision fiel auf, dass der Anteil der relevanten Treffer niedrig war. Selbst der „Testsieger“ Google erreichte nur eine Precision von 0,46. Dies bedeutet, dass nur 46 Prozent der Treffer wirklich relevant sind.

Im Hinblick auf die „schwache“ Precision, bei der alle, sogar wenig relevante Treffer berücksichtigt wurden, betrug die Precision von Google 0,75. Es blieben noch 25 Prozent der in der Top Ten ausgegebenen Treffer, die nicht relevant bzw. schlecht waren.

Bestätigt wird dieser Befund durch eine Untersuchung von Griesbaum, der einen Retrievaltest an „deutschen“ Suchmaschinen (AltaVista.de, Fireball.de, Google.de und Lycos.de) durchführte. Die Ergebnisse zeigten,

„dass etwa 45 Prozent der in den Top 20 ausgegebenen Treffer nicht relevant sind und auch auf kein relevantes Dokument verweisen. Werden diese Werte nicht in Relation zu den anderen, schlechter abscheidenden Suchmaschinen betrachtet, sondern nur in Hinblick auf den Anteil der relevanten Treffer, so ist das Ergebnis insgesamt als schlecht zu bezeichnen“ [Lewa05].

3.2 Weitere Bewertungskriterien

Neben Relevanz wurden im Haupttest die Bewertungskriterien für die Qualität der Treffer aus dem Vortest übernommen: Aktualität, Verständlichkeit, ausreichende Information und Vertrauenswürdigkeit. Im Folgenden wird auf jedes von ihnen einzeln eingegangen.

3.2.1 Verständlichkeit

Bezogen auf die gesamte Treffermenge waren 84 Prozent der Treffer für die Benutzer verständlich (vgl. Abb. 3.6).

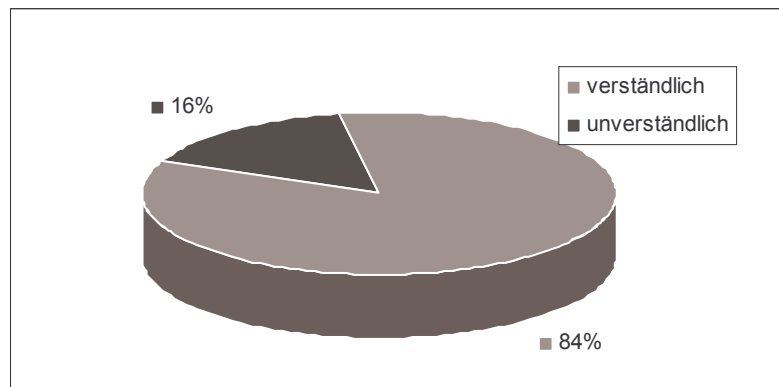


Abb. 3.6 Verständlichkeit der Treffer (n(gesamt)= 2729)

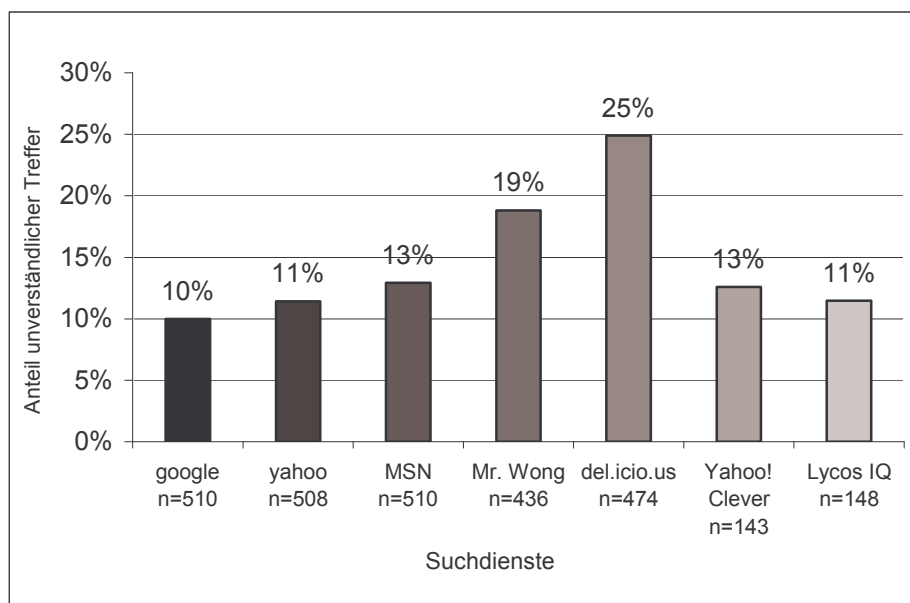


Abb. 3.7 Vergleich der Suchdienste nach Anteil der verständlichen Treffer

Beim Vergleich der Suchdienste nach dem Anteil der Treffer, die für Juroren unverständlich waren (Abb. 3.7), wurde erneut deutlich, dass Social Bookmarking-Dienste verstärkt diese Art Treffer aufweisen. Dabei fiel auf, dass bei del.icio.us 25 Prozent aller Treffer von den Juroren als unverständlich bewertet wurden. Die anderen Suchdienste dagegen hatten einen relativ geringen Anteil an unverständlichen Treffern.

3.2.2 Vertrauenswürdigkeit

Bei der Auswertung der Vertrauenswürdigkeit der Seiten wurden Frage-Antwort-Dienste nicht berücksichtigt, da sie nicht nach diesem Kriterium abgefragt wurden. Bezogen auf die gesamte Treffermenge wurden 49 Prozent der Treffer von den Probanden als vertrauenswürdig und 46 Prozent als nicht vertrauenswürdig eingestuft. Fünf Prozent der Treffer waren, wie bei der Auswertung der anderen Kriterien, irrelevant oder wurden von den Probanden nicht bewertet.

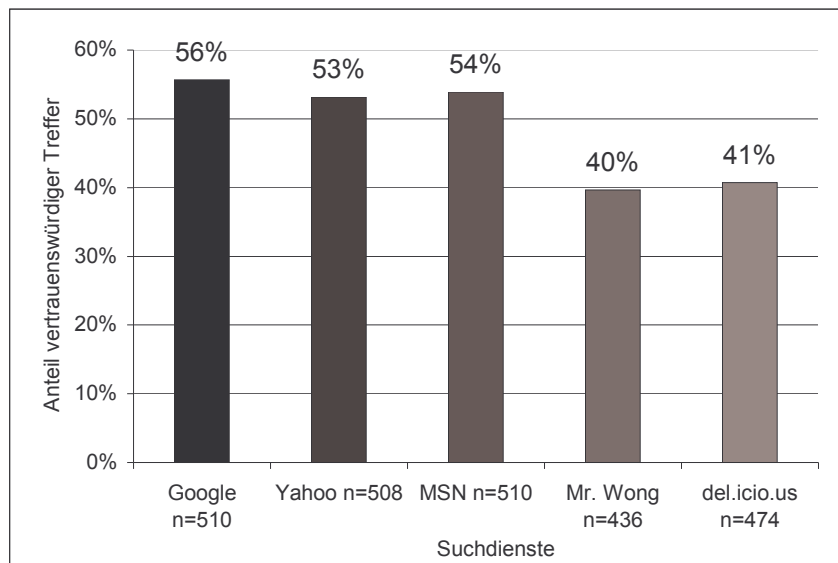


Abb. 3.8 Vergleich der Suchdienste mit Anteil der Treffer, die für Juroren vertrauenswürdig sind (n=2438)

Bei dem Vergleich des Trefferanteils der Suchdienste, die für Juroren vertrauenswürdig sind, führten die algorithmische Suchmaschinen. Die Social Bookmarking-Dienste erreichten nur ca. 40 Prozent der Treffer, die von den Nutzern als vertrauenswürdig beurteilt werden. Dies bedeutet, dass Social Bookmarking-Dienste nicht die von uns erwartete Qualität von Webseiten aufweisen. Wie in der einschlägigen Literatur diskutiert, tendierten auch wir zu der Annahme, dass die von Nutzern ausgewählten Webseiten eine höhere Qualität besitzen, als diejenigen, die von Suchmaschinen gefunden wurden [MaGr07].

3.2.3 Wikipedia-Seiten

Als Nebenprodukt des Relevanztests erschien es interessant, Wikipedia-Seiten zu untersuchen. Das Klischee, Wikipedia-Seiten seien nicht vertrauenswürdig, weil darin beliebig viele Personen Inhalte publizieren können, wurde von den Nutzern anders beurteilt. Obwohl bei Wikipedia-Seiten oft unklar ist, woher die Informationen stammen, wurden sie dennoch von den Nutzern als vertrauenswürdig angesehen.

Insgesamt waren fünf Prozent aller Treffer Wikipedia-Seiten, von denen 63 Prozent als vertrauenswürdig erachtet wurden.

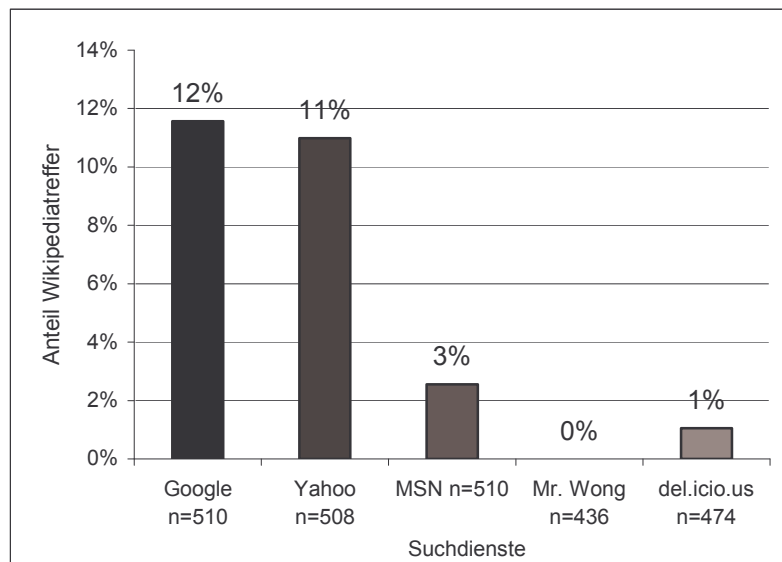


Abb. 3.9 Suchdienste (nur: algorithmische Suchmaschinen und Social-Bookmarking-Dienste (n=2438)) mit den meisten Wikipedia Treffern

Die meisten Wikipedia-Seiten tauchten bei den algorithmischen Suchmaschinen auf. Bei Social Bookmarking-Diensten hingegen wurde sehr wenig auf Wikipedia-Seiten verwiesen.

3.2.4 Ausreichende Information

Das Kriterium „Ausreichende Information“ ist für Frage-Antwort-Dienste nicht anwendbar, weil nur die Fragen dieser Suchdienste untersucht wurden. Das Gesamtergebnis zeigte, dass nur 24 Prozent aller ausgegebenen Treffer ausreichend Informationen boten. Bezogen auf die einzelnen Suchdienste ergab sich das folgende Bild.

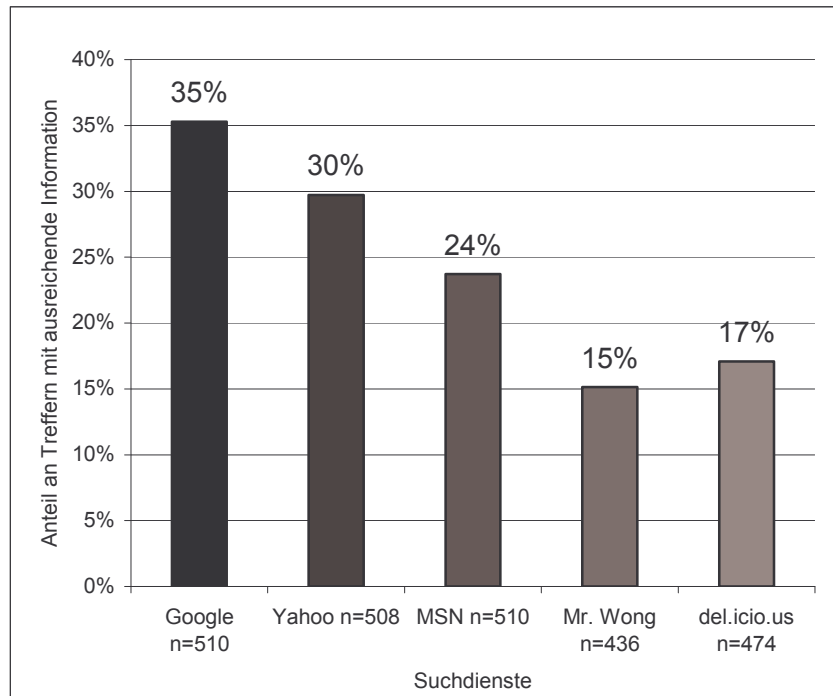


Abb. 3.10 Anteil an Treffern „ausreichende Information“ verteilt nach Suchdiensten

Google erzielte auch hier die höchsten Werte, Yahoo! lag an zweiter Stelle, gefolgt von MSN. Bei den Social Bookmarking-Diensten beinhalteten nur ca. 15 Prozent der ausgegebenen Treffer ausreichende Informationen. Das bedeutet, dass bei den Suchmaschinen häufig bereits beim ersten Treffer alle gesuchten Informationen gefunden werden. Wohingegen der Suchende mit großer Wahrscheinlichkeit bei den Social Bookmarking-Diensten mehrere Treffer ansehen muss, um die gewünschten Informationen zu erhalten.

3.3 Überschneidungen

Unter Überschneidungen versteht man Links aus der Trefferliste eines Suchdienstes, die ein anderer Suchdienst ebenfalls in seiner Trefferliste auflistet. Dabei handelt es sich nicht lediglich um die gleichen URLs, sondern auch um sich minimal unterscheidende Links, die zur gleichen Seite führen. Links, die mehrmals in derselben Trefferliste auftauchen, bezeichnet man als doppelte Treffer.

Da Frage-Antwort-Dienste keine Trefferliste mit Links vorzuweisen haben, sondern hier eine natürlich-sprachige Frage den Treffer ausmacht, tauchen Frage-Antwort-Dienste in diesem Teil der Untersuchung nicht auf.

Von den 2438 Treffern unserer Stichprobe, ohne Frage-Antwort-Dienste, waren 1020 Überschneidungen oder doppelte Treffer. Die meisten Überschneidungen lieferte Google, während die Trefferliste von Mr. Wong die wenigsten Übereinstimmungen mit den Trefferlisten der anderen Suchdienste aufwies. Hierbei

war es interessant zu untersuchen, welche Suchdienste die gleichen Treffer liefern. So stimmten zum Beispiel 34 Prozent der Treffer von Yahoo! und Google überein.

Bei den individuellen Prozentzahlen gab es geringfügige Abweichungen, da die Gesamtmenge der Treffer bei jedem untersuchten Suchdienst unterschiedlich groß war (Abb. 3.11). Diese Unregelmäßigkeiten ergaben sich aus der unterschiedlichen Anzahl von Treffern, die der jeweilige Suchdienst zu einigen der Suchanfragen aufgelistet hatte. Während die Suchmaschinen so gut wie immer mindestens zehn Treffer fanden, kam es bei den Social-Bookmarking-Diensten des Öfteren vor, dass nur fünf bis acht Treffer aufgelistet wurden.

Überschnidungen	Google	Yahoo!	MSN	del.icio.us	Mr. Wong
Google n=510	0,98%	34,06%	28,43%	13,08%	12,84%
Yahoo! n=508	33,92%	1,57%	26,67%	10,34%	10,09%
MSN n=510	28,43%	26,77%	1,96%	7,81%	8,94%
del.icio.us n=474	12,16%	9,65%	7,25%	4,22%	15,14%
Mr. Wong n=436	10,98%	8,66%	7,65%	13,92%	2,06%

Abb 3.11 Anteil der Übereinstimmung der verschiedenen Suchdienste

Wenn die Treffer zusammengerechnet werden, die sowohl von Google als auch von einem oder mehreren anderen Suchdiensten gefunden wurden, ergeben sich 50 Prozent. Das heißt, dass 50 Prozent der Treffer, die andere Suchdienste lieferten, ebenfalls bei Google gefunden werden. Die größte Übereinstimmung findet sich hierbei zwischen Yahoo! und MSN. Yahoo! findet 48 Prozent der Treffer, die auch die anderen Suchdienste aufgelistet haben und bei MSN gibt es eine Übereinstimmung von 40 Prozent. Del.icio.us hingegen teilt sich die meisten Überschneidungen mit Mr. Wong und umgekehrt.

Eine Besonderheit sind die doppelten Treffer, in Abb. 3.11 kursiv dargestellt (siehe auch Abb. 3.14), die sich nicht bei anderen Suchdiensten finden, sondern in der gleichen Trefferliste. Hierbei liegt del.icio.us mit rund vier Prozent eindeutig vorn. Das hat zu der Vermutung geführt, dass hier keine doppelten URLs erkannt und automatisch zusammengeführt werden, wenn ein Nutzer sie hinzufügt. Da hier in der Trefferliste selbst die URLs nicht sichtbar sind, könnte dies bei der Suche schnell zu Frustration bei den Nutzern führen.

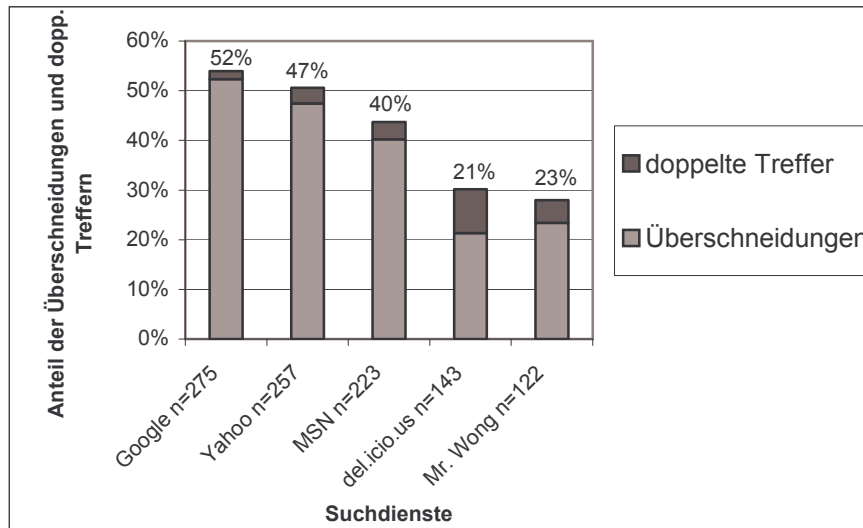


Abb. 3.12 Verteilung der Überschneidungen und doppelten Treffer (n=1020) auf die Suchdienste

Google und Yahoo! wiesen besonders bei Treffern mit „mittlerer“ Precision, aber auch bei den irrelevanten Treffern, Überschneidungen mit anderen Suchdiensten auf. Die Ähnlichkeit der Verteilung dieser beiden Suchdienste ergab sich aus der großen Übereinstimmung ihrer Treffer. Auch MSN zeigte eher Überschneidungen bei Treffern mit „mittlerer“ Precision. Im Gegensatz zu seinen Konkurrenten fanden sich hier jedoch vermehrt Überschneidungen bei weniger relevanten und irrelevanten Treffern.

Die Social Bookmarking-Dienste zeigten ebenfalls Überschneidungen bei relevanten Treffern, doch eine Häufung war eindeutig bei den irrelevanten Treffern zu sehen.

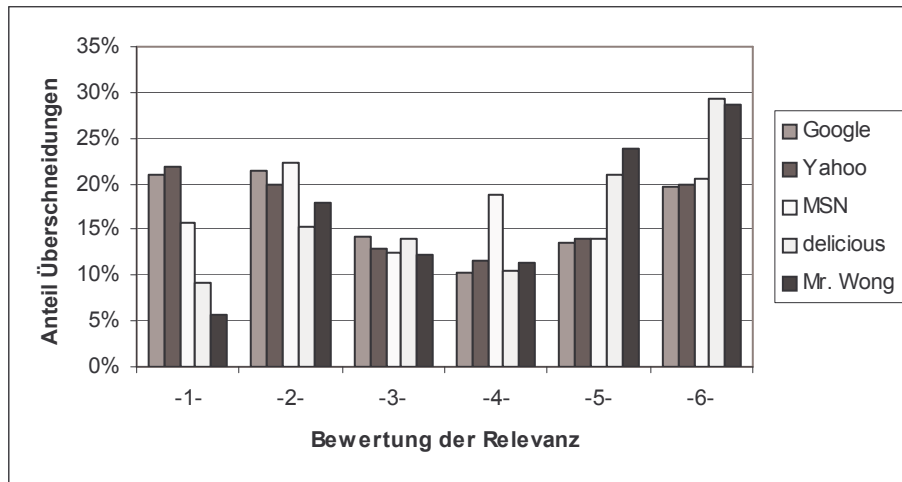


Abb. 3.13 Relevanz der Überschneidungen und doppelten Treffer (n=1020) bei den einzelnen Suchdiensten

Bei allen Suchdiensten fanden sich die meisten Überschneidungen und doppelten Treffer auf den Rängen eins bis fünf, wobei sie sich besonders auf den beiden ersten Plätzen häuften. Bei Mr. Wong zeigte sich eine besondere Konzentrierung auf die ersten beiden Ränge.

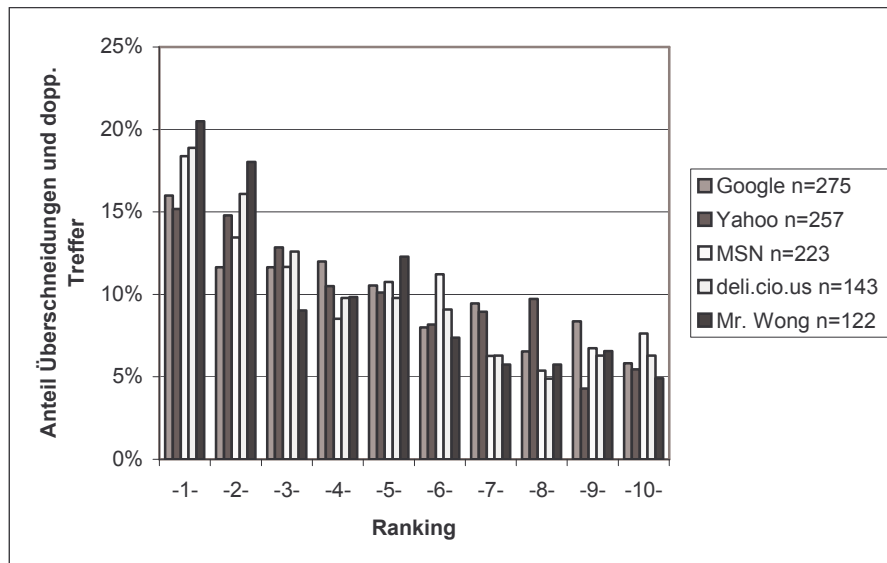


Abb. 3.14 Ranking der Überschneidungen und der doppelten Treffer (n=1020)

Die Verteilung der Überschneidungen und doppelten Treffer auf die Suchanfragen zeigte, dass hier nochmals eine Unterscheidung zwischen den sich überschneidenden Treffern stattfindet. Zum einen ging es um die identischen URLs, die die Suchdienste

auf die gestellten Suchanfragen ausgaben, und zum anderen um die Seiten, auf die diese URLs verwiesen. Bei „Eishockey“ und „Zeitumstellung“ wiesen die meisten doppelten Treffer, im Durchschnitt 3,2, auf die gleiche Seite. Hier waren also bei einer Suche in unterschiedlichen Suchdiensten immer wieder die gleichen Treffer erhältlich. Bei „dd“ hingegen lieferten die meisten Suchdienste unterschiedliche Treffer, was wahrscheinlich an der hohen Mehrdeutigkeit der Suchanfrage lag (siehe auch Anhang).

3.4 Irrelevante Treffer

Abb. 4.15 gibt die “negative” Precision an, also die Precision auf die Relevanzbewertung fünf und sechs. Dabei sticht hervor, dass die Social Bookmarking-Dienste am schlechtesten abschnitten. Auch die Frage-Antwort-Dienste wiesen keine befriedigenden Ergebnisse auf. Eine Ausnahme bildete allerdings der erste Treffer bei Lycos iQ, dessen Wert besser war als der der zweiten Treffer von Yahoo! und MSN. Der für Suchanfragen auf Deutsch am wenigsten geeignete Suchdienst war del.icio.us, dessen Precision-Wert auf die Bewertung fünf und sechs 0,5 nie unterschritt.

Bei den algorithmischen Suchmaschinen zeichnete sich erwartungsgemäß unter den ersten Treffern ein recht guter Wert ab. Diesen guten Wert hielt jedoch nur Google bis zum fünften Treffer. Die anderen Dienste lieferten zum Teil bereits ab dem zweiten Treffer negative Ergebnisse (vgl. Abb. 3.15). Dieses Ergebnis bestätigt die Aussagen der „normalen Precision“. Des Weiteren ist in der Tabelle die Verteilung der nicht relevanten Treffer auf den Plätzen der Rankinglisten angegeben. So wies Google auf Platz eins – über alle Anfragen hinweg – nur zwölf nicht relevante Treffer auf, Mr. Wong hingegen 32.

Ranking	Google Trefferanzahl	Precision?	Yahoo! Trefferanzahl	Precision?	MSN Trefferanzahl	Precision?	Mr. Wong Trefferanzahl	Precision?	del.icio.us Trefferanzahl	Precision?	Yahoo! Clever Trefferanzahl	Precision?	Lycos IQ Trefferanzahl	Precision?
1	12,00	0,24	19,00	0,37	20,00	0,39	32,00	0,63	30,00	0,61	24,00	0,47	19,00	0,37
2	13,00	0,25	22,00	0,43	20,00	0,39	26,00	0,52	32,00	0,63	22,00	0,48	23,00	0,47
3	14,00	0,29	21,00	0,41	18,00	0,35	26,00	0,58	28,00	0,51	21,00	0,46	23,00	0,48
4	17,00	0,31	15,00	0,29	21,00	0,41	21,00	0,47	35,00	0,69				
5	16,00	0,31	23,00	0,45	28,00	0,55	30,00	0,64	31,00	0,61				
6	23,00	0,45	21,00	0,41	26,00	0,51	23,00	0,53	24,00	0,51				
7	24,00	0,47	24,00	0,48	26,00	0,51	27,00	0,66	22,00	0,53				
8	15,00	0,29	23,00	0,45	32,00	0,63	25,00	0,63	29,00	0,66				
9	23,00	0,45	27,00	0,53	30,00	0,59	19,00	0,51	26,00	0,60				
10	23,00	0,45	25,00	0,50	26,00	0,51	29,00	0,78	28,00	0,67				

Abb 3.15 „negative“ Precision

Begründung	Häufigkeit
Suchanfrage	6
Fremdsprache	14
Anderes	18
Subjektive Frage	21
Error	67
Zu wenig Information	73
Kommerziell	564
Inhalt passt nicht	563
Gesamtmenge "irrelevante Treffer"	1332

Tab. 3.16 Begründung für die Bewertung eines Treffers als „irrelevant“

Um einen begründeten Vergleich anzustellen, warum ein Treffer als irrelevant bewertet wurde, war es zunächst nötig, die freien Begründungen in verschiedene Kategorien zu unterteilen (vgl. Abb. 3.16), von denen vier einer näheren Erläuterung bedürfen.

Die Kategorie „kommerziell“ umfasst alle Treffer, die zwar zur Suchanfrage passen, aber rein kommerzielle Angebote sind. Dementsprechend finden sich in der Kategorie „Inhalt passt nicht“ alle Treffer, die nicht zum Thema passen – unabhängig davon, ob es sich um ein kommerzielles Angebot oder eine informationsorientierte Seite zu einem anderen Thema handelt.

Unter „Anderes“ sind die Bemerkungen zusammengefasst, die im einstelligen Bereich lagen, aber zu keiner der anderen Kategorien passten, wie z. B. „nur Links“ (sechs Nennungen) oder „Forum zum Thema“ (drei Nennungen).

„Subjektive Frage“ ist die einzige Kategorie, die nur für die Frage-Antwort-Dienste zählt. Sie erfasst jene Fragen, die auf die Meinung der Community abzielen und somit keine objektive Information generieren..

Die Auswertung der einzelnen Kategorien ergab vor allem bei den Social Bookmarking-Diensten einige Mängel. So entstammten 58,2 Prozent der Seiten, welche nicht geladen werden konnten, einer Anfrage an del.icio.us oder Mr. Wong. Von den 14 Treffern, die aufgrund ihrer Sprache nicht verstanden werden konnten, entstammten 13 del.icio.us, dem einzigen untersuchten Dienst ohne deutsche Lokalisierung. Ein weiterer Treffer wurde von MSN gefunden.

In den Kategorien „kommerziell“ und „Inhalt passt nicht“ wurden die Prozente der Kategorie aus der Gesamtmenge aller gefundenen Treffer innerhalb eines Suchdienstes errechnet. Hierbei fiel auf, dass bei allen algorithmischen Suchmaschinen die kommerziellen Seiten deutlich mehr Raum einnahmen als die unpassenden. (Google: 18,24 Prozent kommerziell, 13,14 Prozent unpassend; Yahoo!: 24,07 Prozent kommerziell, 15,88 Prozent unpassend; MSN: 28,04 Prozent kommerziell, 14,51 Prozent unpassend).

Dieses Ergebnis ist wahrscheinlich dem Versuch zuzuschreiben, möglichst alle interessanten Seiten zu finden, da die Maschinen nicht in der Lage sind, eine informationsorientierte Anfrage zu erkennen. Dem gegenüber stehen die Social Bookmarking-Dienste, deren Treffer eher unpassend als kommerziell sind. Dabei erscheinen die Zahlen für kommerzielle Treffer immer noch überraschend hoch, wenn man bedenkt, dass der Index dieser Dienste von einer Community aufgebaut wird

(Mr. Wong: 23,39 Prozent kommerziell, 28,21 Prozent unpassend; del.icio.us: 21,52 Prozent kommerziell, 28,27 Prozent unpassend).

Da die Frage-Antwort-Dienste bisher frei von kommerziellen Treffern sind, wurden hier nur die Werte für unpassende Treffer verglichen. Das Ergebnis dieses Vergleichs (Yahoo! Clever: 30,07 Prozent; Lycos iQ: 27,7 Prozent) zeigte aber wiederholt, dass das Ranking der Fragen verbessert werden müsste.

Beim Vergleich der Trefferanteile innerhalb der unpassenden Treffer zeigte sich, dass die algorithmischen Suchmaschinen im Mittelfeld lagen (Google: 12 Prozent, Yahoo!: 14 Prozent, MSN: 13 Prozent), während die Frage-Antwort-Dienste am wenigsten unpassende Treffer lieferten (Yahoo! Clever: 8 Prozent; Lycos iQ: 7 Prozent). Überraschend war die Tatsache, dass die Social Bookmarking-Dienste – trotz der geringeren Gesamttreffermenge – die meisten unpassenden Treffer fanden (Mr. Wong: 22 Prozent; del.icio.us: 24 Prozent) (vgl. Abb. 3.17).

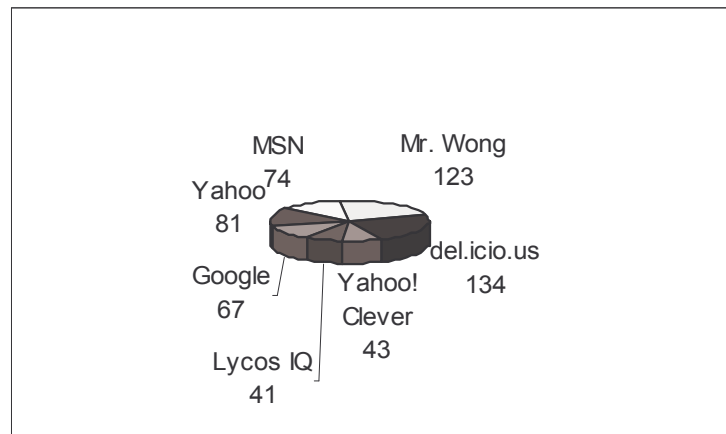


Abb. 3.17 Anteil an der Gesamtmenge „unpassender Treffer“ (n=563) verteilt nach Suchdienst

Beim Vergleich der kommerziellen Treffer pro Suchdienst lagen die Social Bookmarking-Dienste (je 18 Prozent) nur einen Prozentpunkt hinter Google (17 Prozent) (vgl. Abb. 4.14). Die anderen klassischen Suchmaschinen lagen bei 22 und 25 Prozent und produzierten somit fast die Hälfte der gesamten kommerziellen Treffer.

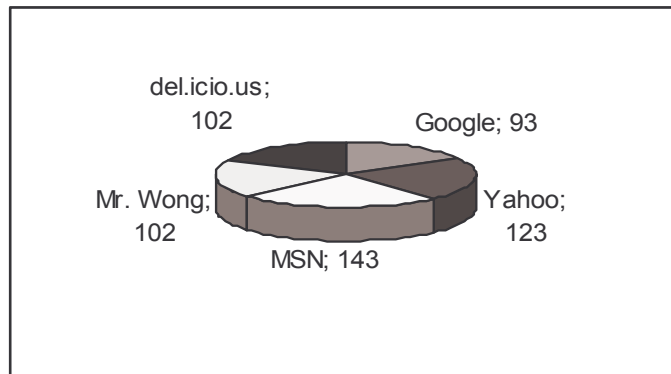


Abb. 3.18 Anteil an der Gesamtmenge „kommerzieller Treffer“ (n=564) - verteilt nach Suchdienst

Bei den als subjektiv bewerteten Fragen schniedete Yahoo! Clever mit 6,2 Prozent besser ab als Lycos IQ mit 8,1 Prozent. Der direkte Vergleich der absoluten Zahlen neun und zwölf zeigt, dass der Anteil von subjektiven Fragen bei beiden Diensten ähnlich war.

4. FAZIT

Aus den Ergebnissen wird ersichtlich, dass im deutschsprachigen Raum Suchmaschinen, hierbei allen voran Google, bei der Bereitstellung relevanter Treffer noch immer am besten abschneiden.

Die irrelevantesten Ergebnisse erzielten überraschenderweise die Social Bookmarking-Dienste. Die Erwartung, dass die von Nutzern generierten Linklisten eine höhere Precision als Frage-Antwort-Dienste aufweisen würden, erwies sich als falsch. Wir nahmen an, dass gute Links von vielen Personen gespeichert werden und somit eine hohe Position in der Trefferliste einnehmen würden [MaGr07]. Diese Vermutung hat sich durch die Untersuchung nicht bestätigt.

Stattdessen gab es besonders bei Social Bookmarking-Diensten sehr viele tote Links und doppelte Treffer in ein und derselben Trefferliste. Hier fehlt eine vernünftige Software, die den Bestand wartet. Der Nutzer scheint dazu nicht fähig oder hat kein Interesse daran. Somit ist unser Eindruck von diesen Diensten, dass sie zwar zur persönlichen Verwaltung von Lesezeichen und zum Austausch von Links innerhalb einer Gruppe sehr gut geeignet, aber als Suchplattform vorerst nicht weiter zu gebrauchen sind.

Entgegen den Ergebnissen bisheriger Studien zeigte Google große Übereinstimmung mit den anderen Suchdiensten. Es läßt sich behaupten, dass eine große Wahrscheinlichkeit besteht, dass Google den relevanten Link, den Yahoo! oder MSN einer suchenden Person ausgibt, ebenfalls findet.

Wie frühere Studien schon gezeigt haben, offeriert Google zwar insgesamt die relevantesten Links, liefert aber, für sich gesehen, gerade einmal 36 Prozent stark relevante Treffer. 35 Prozent der Treffer hingegen besitzen überhaupt keine Relevanz [GRB02] [Grie04] [Véro06]. Das ist kein befriedigendes Ergebnis.

Bei allen anderen Suchdiensten gab es mehr irrelevante als stark relevante Treffer. Bei der Einbindung der Fragen aus den Frage-Antwort-Diensten empfehlen wir, sie bei einer Google-Trefferliste auf den sechsten Rang zu platzieren, da ab hier bei Google die Precision stark abnimmt und somit die Frage aus dem Frage-Antwort-Dienst einen höheren Precision-Wert erreicht als der sich dort befindliche Treffer. Bei einer Trefferliste von Yahoo! würde sich das Einbinden nach dem dritten Treffer lohnen.

Weiterhin empfehlen wir Lycos, bei Lycos iQ ein besseres Ranking einzuführen. Dadurch würde sich die Precision vermutlich erhöhen. Bei unserem Test war es nicht ersichtlich, nach welchen Kriterien die Fragen geordnet wurden. Auch wäre eine Verbesserung der Suchfunktionen notwendig. Phrasenerkennung und eine automatische Berichtigung der Schreibweise sollten zum Standard gehören. Ein Ranking der Antworten, z. B. nach Top-Bewertung, würde bei einer zukünftigen Einbettung von Lycos iQ in eine Trefferliste ebenfalls hilfreich sein.

Verwendete Literatur

- [GRB02] Griesbaum, Joachim/Rittberger, Marc/Bekavac, Bernard (2002): Deutsche Suchmaschinen im Vergleich: Alta Vista.de, Fireball.de, Google.de und Lycos.de. In: Womser-Hacker, Christa/Wolff, Christian/Hammwöhner, Reinhard (Hrsg.): Information und Mobilität: Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposium für Informationswissenschaft (ISI 2002). UVK: Konstanz. S. 201–223. Onlinedokument: <http://www.inf-wiss.uni-konstanz.de/infwiss/download/isi2002/cc-isi2002-art14.pdf> [Abruf am 16.01.2008]
- [Grie00] Griesbaum, Joachim (2000): Evaluierung hybrider Suchsysteme im WWW. Diplomarbeit Universität Konstanz. Onlinedokument: http://www.inf.uni-konstanz.de/Prozent7Egriesbau/files/evaluierung_hybrider_suchsysteme_im_www.pdf [Abruf am 16.01.2008]
- [Grie04] Griesbaum, Joachim (2004): Evaluation of three German search engines: Altavista.de, Google.de and Lycos.de. In: Information Research Bd. 9, H. 4, S. 189. Onlinedokument: <http://informationr.net/ir/9-4/paper189.html> [Abruf am 28.12.2007]
- [LeHö07] Lewandowski, Dirk/Höchstötter, Nadine (2007): Qualitätsmessung bei Suchmaschinen: System- und nutzerbezogene Evaluationsmaße. In: Informatik Spektrum, Bd. 30, H. 3, S. 1–11. Onlinedokument: http://www.durchdenken.de/lewandowski/doc/IS_2007_Preprint.pdf [Abruf am 16.01.2008]
- [Lewa05] Lewandowski, Dirk (2005): Web Information Retrieval: Technologien zur Informationssuche im Internet. Onlinedokument: http://www.durchdenken.de/lewandowski/web-ir/?25_Arten_von_Suchanfragen.html. [Abruf am 06.12.2007]
- [Lewa06] Lewandowski, Dirk (2006): Themen und Typen der Suchanfragen an deutsche Web-Suchmaschinen. Onlinedokument: <http://www.durchdenken.de/lewandowski/doc/mkwi2006.pdf>. [Abruf am 26.12.2007]
- [MaGr07] Maaß, Christian/Gräfe, Gernot (2007): Alternative Suchdienste: sieben Thesen zur Bedeutung des „Social Bookmarking“ In: Praxis der Wirtschaftsinformatik HMW (in Drucklegung) S. 1–12.
- [Mizz97] Mizzaro, Stefano (1997): Relevance: the whole story. In: Journal of the American Society for Information Science, Bd. 48, S. 810–832.
- [Sara96] Saracevic, Tefko (1996): Relevance reconsidered. In: Information Science: Integration in Perspectives. Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2). Kopenhagen. S. 201–218.

- [Stoc07] Stock, Wolfgang G. (2007): Infomation Retrieval: Informationen suchen und finden. Oldenbourg Wissenschaftsverlag: München.
- [StSt00] Stock, Metchild/Stock, Wolfgang G. (2000): Internet – Suchwerkzeuge im Vergleich: Test 1: Retrievaltest mit Known Item Searches. In. Password, Bd. 15, H. 11, S. 23-31. Onlinedokument: http://philfak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1078739217password_1.pdf [Abruf am 16.01.2008]
- [Véro06] Véronis, Jean (2006): A comparative study of six search engines. Onlinedokument: <http://sites.univ-provence.fr/veronis/pdf/2006-comparative-study.pdf> [Abruf am 28.12.2007]

Weiterführende Literatur

1. Womser-Hacker, Chista (1989), Der PADOK-Retrievaltest. Zur Methode und Verwendung statistischer Verfahren bei der Bewertung von Information-Retrieval-Systemen. Hildesheim et al. (Dissertation Universität Regensburg, Linguistische Informationswissenschaft)
2. Machill, Marcel/Welp, Carsten (Hrsg.) (2003): Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen. Bertelsmann Stifung: Gütersloh.

Untersuchung der Qualität der Antworten bei Lycos iQ und deren Einbindung in die algorithmische Suche

KONSTANTIN GURATZSCH

kguratzsch@web.de

SONJA VON MACH

SovMa@gmx.de

JESSICA OTTE

jessicaotte@gmx.de

JOËL PIENTOK

joelpientok@gmx.de

Der Bericht befasst sich mit der Untersuchung, der Qualität der bei Lycos iQ abgegebenen Antworten. Die Untersuchung wurde anhand einer Antworten-Analyse durchgeführt, wofür ein auf dieses Ziel zugeschnittener Kriterienkatalog erstellt wurde. Die Ergebnisse des Berichts zeigen, dass die Qualität der Antworten bei Lycos iQ, mit zu 80 Prozent richtigen Rückmeldungen als hoch bezeichnet werden kann. Als weiteres Ergebnis kann die Empfehlung zur Einbindung in die Websuche insofern ausgesprochen werden, als dass die höchste Bewertung, die eine Antwort erreichen kann, auch zu 92 Prozent die Richtigkeit einer Antwort garantiert. Demnach sollten lediglich hoch bewertete Antworten eingebunden werden.

1. EINLEITUNG

Da Suchmaschinen oft nicht die präzise Antwort auf eine Suchanfrage bieten, erfreuen sich Frage-Antwort-Dienste immer höherer Beliebtheit. Hier kann eine Frage genau formuliert werden, ohne nur einige Schlüsselworte in das Suchfeld einzugeben, was oft zu ungenauen Ergebnissen führt.

So konnten sich allein auf dem deutschen Internetmarkt neben Lycos iQ die Frage-Antwort-Dienste Yahoo! Clever, Wer-weiss-was, Gutefrage.net und Fragr etablieren. Bisher noch unerforscht geblieben ist aber die Frage, wie die Qualität der Antworten der freiwilligen Ratgeber eigentlich ist. Diese Arbeit greift sich den Frage-Antwort-Dienst Lycos iQ heraus und konzentriert sich auf die Untersuchung dieser Fragestellung. Dabei stehen drei Forschungsfragen im Vordergrund:

1. Feststellung der Qualität der Antworten von Lycos iQ
2. Herausarbeiten, ob der Expertenstatus mit qualitativ hochwertigen Antworten korreliert

3. Qualitätsmerkmale von Antworten erarbeiten, um die Einbindung in die algorithmische Suche zu ermöglichen

- Forschungsfrage 1: Welche Qualität haben die Antworten auf Lycos iQ?

Es wurden dazu 200 Fragen herausgegriffen und deren Antworten untersucht (insgesamt 686 Antworten). Für die Untersuchung stand uns ein Team von 20 Studenten zur Verfügung. Um mit möglichst klar strukturierten Daten zu arbeiten, mussten die Fragen im Frage-Antwort-Dienst Lycos iQ in Themengruppen eingeteilt werden. Als Instrument für die Untersuchung der Qualität der Antworten sowie der Forschungsfragen 2 und 3 wurde von uns ein Kriterienkatalog entwickelt. Damit sollte jede Antwort nach 58 Kriterien geprüft werden, um möglichst viele Informationen zu jeder einzelnen Antwort zu erhalten.

- Forschungsfrage 2: Korreliert der Expertenstatus mit qualitativ hochwertigen Antworten?

Der Expertenstatus der Antwortenden kommt durch ein Bewertungssystem zustande. Der Fragensteller kann die Qualität der Antworten bewerten. Ein Punktesystem belohnt die Ratgeber für ihre Hilfe. So kann derjenige, der eine Frage bei Lycos iQ beantwortet, in der Expertenhierarchie vom Einsteiger bis zum Einstein aufsteigen. Die zweite Forschungsfrage beleuchtet also die Frage, inwieweit der Expertenstatus der freiwilligen Ratgeber sich auf die Qualität der Antworten auswirkt.

- Forschungsfrage 3: Lassen sich Qualitätsmerkmale erarbeiten, anhand derer die Einbindung in die algorithmische Suche möglich wäre?

Mit der dritten Forschungsfrage ist folgende Möglichkeit gemeint: Zeichnet sich ab, dass ein bestimmter Expertenstatus Garant für qualitativ hochwertige Antworten ist, dann könnte dieser Expertenstatus mit seinen Antworten automatisch in das algorithmische Suchergebnis eingebunden werden/ dann könnten die Antworten von Nutzern mit solchem Expertenstatus automatisch....Eine weitere Möglichkeit wäre aber auch, dass eine bestimmte Bewertungskategorie richtige Antworten liefert und somit alle Antworten, die diese Bewertung erfahren, in die Websuche eingebunden werden könnten.

2. METHODISCHE VORGEHENSWEISE

Für die Beantwortung der drei Forschungsfragen musste zunächst dafür gesorgt werden, dass die zu untersuchenden Antworten nicht willkürlich aus Lycos iQ herausgegriffen wurden, sondern inhaltlich aus den wichtigsten unterschiedlichen Themengebiete von Lycos iQ stammten. Dazu wurden im ersten Schritt zehn Themengruppen gebildet.

2.1 Einteilung der Antworten in Themengruppen

Da der Frage-Antwort-Dienst von Lycos iQ über 220.000 Themen aufführt, wurden diese auf eine überschaubare Zahl von zehn Themengruppen reduziert und zusammengefasst (vgl. Abb. 4.1). Dabei konzentrierte sich die Zusammenfassung nur auf die Themen, die zu dem Zeitpunkt der Untersuchung mindestens 1000 Mal von Usern verwendet worden waren (vgl. Abb. 4.2).

1. Musik
2. Computer & Technisches
3. Wissenschaften
4. Medizin/Gesundheit
5. Freizeit
6. Filme
7. Sprachen/Länder
8. Tiere & Natur
9. Beziehung / Liebe / Zwischenmenschliches
10. Gesellschaft

Abb. 4.1 Themeneinteilung der Antworten in Lycos iQ

```
Computer & Technisches (Anzahl der Fragen: 60.093)
2.computer(13788)
3.internet(11617)
4.lycos iq(8442)
6.PC(7999)
33.software(2973)
106.Games(1029)
107.hardware(1013)
87.lycos(1292)
97.online(1117)
68.Windows(1620)
83.download(1331)
89.homepage(1238)
29.handy(3416)
31.technik(3218)
```

Abb. 4.2 Beispiel einer zusammengefassten Themengruppe (Computer & Technisches)

2.2 Erstellung des Kriterienkatalogs

Der Kriterienkatalog ist ein Fragekatalog, mit dem jede herausgegriffene Antwort nach 58 Kriterien abgefragt wird. Mittels des Kriterienkatalogs sollten die Antworten von Lycos iQ hinsichtlich der drei Forschungsfragen (vgl. Kap. 1) genau analysiert werden. Der Kriterienkatalog musste deshalb so konzipiert werden, dass alle möglichen Fragentypen (Meinungsfragen, Faktenfragen und Erfahrungsfragen) und Antwortentypen (Top, hilfreich und gut bewertet) erfasst werden konnten.

Für die Erstellung des Kriterienkatalogs wurde daher für die Einteilung der Fragen und die Einteilung der Antworten zunächst ein Analyseraster geschaffen.

2.2.1 Einteilung der Fragen

Im ersten Entwurf des Kriterienkatalogs wurden die Fragen in Meinungsfragen, Erfahrungsfragen und Faktenfragen eingeteilt. Während sich die Faktenfragen auf Tatsachen beziehen, die objektiv für einen Beweis zugänglich sind, sind die Meinungs- und Erfahrungsfragen grundsätzlich subjektiv geprägt. Die Abgrenzung zwischen Meinungs- und Erfahrungsfrage war schwer zu ziehen – zunächst wurde diese Abgrenzung mit der Frage nach dem „Kern“ der Fragestellung gezogen. Beruhte die Frage im „Kern“ auf der persönlichen Erfahrung des Ratgebers mit dem Gegenstand der Frage, so handelte es sich um eine Erfahrungsfrage. Eine Meinungsfrage hingegen lag dann vor, wenn sich die Frage im „Kern“ auf die Stellungnahme bzw. Einschätzung des Ratgebers bezieht.

Nach Erstellung des ersten Entwurfs des Kriterienkatalogs zeigten sich in dem ersten Praxistest (Pretest) deutliche Abgrenzungsschwierigkeiten hinsichtlich der Erfahrungs- und Meinungsfragen. Eine schwierig einzuordnende Abgrenzungsfrage war zum Beispiel: „Was haltet Ihr vom neuen Handy w850i von Sony?“ Diese wurden deshalb zu einem gemeinsamen Fragetyp, der Erfahrungsfrage, zusammengelegt.

Mit der Themeneinteilung sollte ermöglicht werden, aus jedem Themengebiet eine bestimmte Anzahl von Antworten je nach Fragenzugehörigkeit exemplarisch auszuwerten. So sollte das Verhältnis von inhaltlichen Themen und Fragentypen (Fakten- oder Erfahrungsfrage) eingehalten werden, um repräsentative Daten analysieren zu können.

2.2.2 Einteilung der Antworten

Zum Verständnis der Einteilung der Antworten in unterschiedliche Rubriken ist es sinnvoll, die Vorgehensweise bei dem Frage-Antwort-Dienst Lycos iQ darzustellen. Jeder User von Lycos iQ kann eine Frage stellen. Diese Frage kann dann von jedem Lycos iQ User beantwortet werden. Auf eine Frage kann es deshalb mehrere Antworten geben. Der Fragesteller kann Bonuspunkte für seine Frage aussetzen und die Ratgeber für ihre Antworten bewerten. Durch diese Bewertung werden anschließend die ausgesetzten Bonuspunkte verteilt (ein Gastnutzer, der noch nicht in vollem Umfang bei Lycos iQ registriert ist, kann keine Bonuspunkte aussetzen).

Um die Qualität der Antworten bei Lycos iQ messen zu können, mussten wir alle denkbaren Bewertungsmöglichkeiten in dem Kriterienkatalog berücksichtigen. Abbildung 4.2 zeigt beispielhaft eine Frage aus dem Bereich Physik, anhand deren Antwort die unterschiedlichen Bewertungsmöglichkeiten, die im Kriterienkatalog berücksichtigt wurden, gezeigt werden:



iq (Rang: Grace Hopper)

Durch welche Faktoren wird die Aerodynamik eines Gegenstandes bestimmt?



1

Bonuspunkte: 10

Gestellt in Physik, Aerodynamik, Flugeigenschaften 15.01.2008 10:17

geschlossen am: 18.01.2008 10:17

[Gute Frage!](#)



2




Rang: Professor (4.380) | Physik (600) ★

99 Minuten nachdem die Frage gestellt worden ist (15.01.2008 11:56)

Unter aerodynamischen Eigenschaften wäre zu unterscheiden der Strömungswiderstand und der Auftrieb.

Auftrieb wird dadurch erzeugt, dass man ein Profil so gestaltet, dass sich an der Oberseite eine höhere Strömungsgeschwindigkeit einstellt als an der Unterseite.

Der Strömungswiderstand wird durch Verwirbelungen erzeugt, die Energie verschlingen. Für geringen Luftwiderstand muss man Profile so gestalten, dass sie entweder wenig Wirbel verursachen oder diese Wirbel sich von der Oberfläche ablösen und sich innerhalb einer "Wirbelschlepe" totlaufen, ohne dabei in Kontakt mit dem Gegenstand zu bleiben. Auf Oberflächen kann eine Wirbelschicht den Widerstand sogar senken, wie an Golfbällen erkennbar ist, deren große Flugweite mit den Dellen in der Oberfläche zu tun hat.

Bewertung des Fragestellers:  3 Top	Gute Antwort meinen:  4 Als gute Antwort bewerten	Kommentare zur Antwort:  0 Kommentar abgeben
--	--	---

Legende:

- 1 = Fragesteller vergibt zehn mögliche Bonuspunkte für eine gute Beantwortung der Frage
- 2 = Rang des Antwortenden (Professor)
- 3 = Vom Fragesteller Top bewertete Antwort (neben einer Top Bewertung kann der Fragesteller die Antwort alternativ auch mit „nicht hilfreich“ oder „hilfreich“ bewerten)
- 4 = Von anderen Usern als gut bewertete Antwort

Abb. 4.2 Unterschiedliche Bewertungsmöglichkeiten bei Lycos iQ

Angeichts der unterschiedlichen Bewertungsmöglichkeiten des Fragestellers einerseits und der Bewertung durch andere User andererseits ergab sich für den Kriterienkatalog folgende Ordnung:

1. Untersuchung derjenigen richtigen Antworten, die nicht als hilfreich, Top oder gut bewertet wurden

2. Untersuchung der Top Antworten
3. Untersuchung der als hilfreich bewerteten Antworten
4. Untersuchung der von anderen Usern als gut bewerteten Antworten.

2.2.3 Inhalt des Kriterienkatalogs

Unter dem Blickwinkel der Untersuchungsziele mussten für den Kriterienkatalog Fragen formuliert werden, die möglichst genaue Aussagen über die Qualität der Antworten in Korrelation mit dem Expertenstatus liefern sollten. Insgesamt wurden jeweils 58 Kriterien zur Auswertung der Fakten- und Erfahrungsfragen formuliert, die größtenteils identisch waren. Im Wesentlichen ging es darum, so viele Informationen wie möglich zu den Antworten, den Bewertungen, den Themengebieten, der Schnelligkeit der Beantwortung und der Unterstützung der Antwort durch einen Link abzufragen, um ein möglichst umfassendes Bild der Qualität der Antwort zu erhalten.

Allgemeine Fragen

Inhaltlich wurden im Kriterienkatalog zunächst allgemeine Fragen gestellt, die für die statistische Auswertung von Belang waren, z.B.: „Aus welchem Themengebiet kommt diese Frage?“, „Wie viele Antworten wurden gegeben?“, „Wie viele Bonuspunkte wurden von dem Fragenden ausgesetzt?“ und „Wie viele Antworten wurden als Top/hilfreich bewertet?“.

Eine Besonderheit ergab sich bei der Frage nach der Richtigkeit der Antworten („Wie viele Antworten sind richtig bzw. falsch?“). Hier mussten die Eigenheiten von Faktenfragen und Erfahrungsfragen berücksichtigt werden: Während Antworten auf Faktenfragen nachprüfbar sind und damit eindeutig richtig oder falsch sind, können Erfahrungsfragen aufgrund ihrer Subjektivität nicht als richtig oder falsch geschätzt werden. Als Kompromisslösung wurde bei Erfahrungsfragen deswegen einfach gefragt, ob sich die Antwort auf die Frage bezieht oder ob sie am Thema der Frage vorbeigeht („Wie viele Antworten beziehen sich wirklich auf die Frage?“, „Wie viele der Antworten beziehen sich nicht auf die Frage bzw. sind unverständlich?“).

Da Antworten oft nicht nur durch Fließtext, sondern auch durch einen Link gegeben bzw. unterstützt werden, bildeten sich hinsichtlich der Beurteilung der Seriosität der Links drei Abstufungen:

1. Top Quellen	Nachrichtenseiten, amtliche Seiten Beispiel: Bei der Frage nach der Höhe des Eiffelturms wäre eine Top Quelle die offizielle Seite des Eiffelturms
2. Offizielle Quellen	Uni-Seiten, Homepages von Organisationen Beispiel: Organisationen, die sich unter anderem mit dem Eiffelturm beschäftigen
3. Inoffizielle Quellen	Wikis, Blogs, private Homepages Beispiel: Reiseberichte Paris

Abb. 4.4 Einteilung der Links in Kategorien

Der Bearbeiter des Kriterienkatalogs musste angeben, aus welcher der angegebenen Quellen der in der Antwort angegebene Link stammte. Problematisch erschien die Frage, wie der anscheinend oft genutzte Wikipedia- Link in dieser oben aufgezeigten Einstufung eingeordnet werden sollte. Aufgrund der Strittigkeit über die Seriosität von Wikipedia wurde in dem nach dem Pretest überarbeiteten Kriterienkatalog dafür eine eigene Rubrik bzw. Frage gebildet (vgl. Frage 47 in Abb. 4.5). Ob und inwieweit Antworten nur mit Hilfe von Wikipedia-Links oder Wikipedia Auszügen gegeben werden, ist nicht nur für die Beurteilung der Qualität der Antworten wichtig, sondern auch für die Frage, wie eigenständig die freiwilligen Ratgeber Fragen beantworten.

Besondere Fragen

Nach diesen einleitenden allgemeinen Fragen wurden nun die Aspekte der einzelnen Antworten abgefragt. Ziel war es, möglichst viele Informationen zu den einzelnen Antworten innerhalb der unterschiedlichen Bewertungsgruppen zu erhalten, z.B.: „Welchen Rang hat der Herausgeber der richtigen Antwort?“, „Zu was für einem Prozentsatz hat der Herausgeber der richtigen Antwort bereits hilfreiche Antworten gegeben?“, „Hat der Herausgeber der richtigen Antwort in seinem Themengebiet geantwortet?“, „Ist die richtige Antwort ein Fließtext?“, „Ist die richtige Antwort ausschließlich ein Link?“.

Abb. 4.5 zeigt exemplarisch eine Bewertungsgruppe (hilfreich bewertete Antworten) und verdeutlicht welche Fragen gestellt wurden, um die Qualität einer hilfreich bewerteten Antwort zu bestimmen.

Untersuchung der hilfreichen bewerteten Antworten

38. Ist die als hilfreich bewertete Antwort richtig?
39. Welchen Rang hat der Herausgeber der als hilfreich bewerteten Antwort?
40. Zu wie viel Prozent hat der Herausgeber der als hilfreich bewerteten Antwort bereits richtige Antworten gegeben?
41. Hat der Herausgeber der hilfreichen Antwort in seinem Themengebiet geantwortet?
42. Nach wie vielen Minuten hat der Herausgeber der hilfreichen Antwort geantwortet?
43. Ist die als hilfreich bewertete Antwort ein Fließtext?
44. Ist die als hilfreich bewertete Antwort ausschließlich ein Link?
45. Wird die als hilfreich bewertete Antwort durch die Angabe eines Links unterstützt?
46. Aus welcher der von uns angegebenen Quelle stammt der Link?
47. Ist die als hilfreich bewertete Antwort durch einen Wikipedia-Link beantwortet worden?
48. Wie oft wurde die als hilfreich bewertete Antwort von anderen Usern als gut bewertet?

Abb. 4.5 Ausschnitt der Faktenfragen des Kriterienkatalogs „Untersuchung der hilfreichen Antworten“

3. AUSWERTUNG DER FRAGENKATALOGE

3.1 Zuweisung der Themengebiete

Nach der Fertigstellung der Kriterienkataloge begann die Untersuchung. Zur Verfügung stand ein Team von insgesamt 20 Personen. Davon waren zwölf Personen weiblich und acht Personen männlich. Jede Person bekam eines der zuvor zusammengestellten Themengebiete zugewiesen. Immer zwei Personen bekamen das gleiche Themengebiet zugeteilt, wobei sich der Aufgabenbereich dieser beiden Personen dadurch unterschied, dass eine Person die Faktenfragen und die andere Person die Erfahrungsfragen eines Themengebietes bekam. So wurde jedes der zuvor vergebenen Themengebiete von zwei Personen bearbeitet. Auf diese Weise wurden die Antworten von insgesamt 100 Faktenfragen und 100 Erfahrungsfragen untersucht. Die Zuweisung der Themengebiete zu den entsprechenden Teammitgliedern folgte also nicht dem System der Präferenzvergabe, sondern war unabhängig vom Wissen und Interesse der Person.

3.2 Probleme nach der Themenzuweisung

Trotz gründlicher Recherche, mehrfachen Überarbeitens der Kataloge und einem ausführlichen Pretest gab es in der Untersuchungs-Phase Unklarheiten bezüglich der Anwendung der Kriterienkataloge.

Jedes der zusammengefassten Themengebiete (vgl. Kap. 2.1) bestand aus mehreren einzelnen Unterthemen. So konnte es beispielsweise sein, dass das Themengebiet Computer aufgrund des breiten Spektrums an Themen auf Lycos iQ,

über sehr viele Unterthemen verfügte. Ein Problem für die Probanden, welche die Fragenkataloge ausgefüllt haben, war das Bearbeiten der Fragenkataloge in Anbetracht der Fülle von Unterthemen. Da jeder Proband lediglich zehn Fragen bearbeiten sollte, war es schwierig, in nur zehn zu bearbeitenden Fragen jedes der Unterthemen, in welches sein Themengebiet involviert war, zu berücksichtigen.

Hier galt es, möglichst repräsentative Fragen eines Themengebietes herauszusuchen und zu untersuchen. Enthielt ein Themengebiet mehr als zehn Unterthemen, so war es sinnvoll, zur Untersuchung Fragen und deren Antworten herauszufiltern, welche zwei Themen umfassten. Ein Beispiel einer solchen Kombinations-Frage wäre: „Ich möchte möglichst viel kulturelle Events besuchen aber nicht so viel Geld dabei ausgeben. Bitte gebt mir Ratschläge.“ Diese Frage kommt aus dem Themengebiet „Gesellschaft“. Die Antworten dieser Frage würden die Unterthemen „Geld“ und „Kultur“ berühren.

Ein weiterer „Stolperstein“ war die Bewertungen der Antworten durch den Fragensteller. Hier gibt es die folgenden drei Möglichkeiten: „nicht hilfreich“, „hilfreich“ oder „Top“. Diese drei Bewertungen setzen ein aktives Handeln des Fragestellers voraus. Dieser muss entscheiden, inwieweit ihm persönlich die Antwort geholfen hat. Ein weiterer Status einer Antwort kann „nicht bewertet“ sein. Das Problem der Teammitglieder basierte auf der Tatsache, dass die meisten Antworten mit einer der drei Bewertungen und kaum mit dem Status „nicht bewertet“ versehen waren. Bei der Untersuchung der Antworten anhand des Faktenkataloges kam es an der Stelle, an der die „nicht bewerteten“ Antworten angesprochen werden, regelmäßig zu Fragen bezüglich der Existenz des Status „nicht bewertet“.

Man kann sagen, dass der Grund für das seltene Vorkommen einer nicht bewerteten Antwort in dem Grundgedanken bzw. dem „Geiste“ einer Community liegt. Eine Community lebt von aktiven Mitgliedern. Lycos iQ lebt von den Fragestellern ebenso wie von den Antwortenden. Der Reiz des Antwortgebens besteht in der Anerkennung durch andere User sowie in den zu erhaltenden Punkten bei einer entsprechend bewerteten Antwort. Würden die Antwortenden keine Resonanz von den Fragestellern bekommen, so würde der Reiz des Antwortengebens wegfallen. Die Folge wäre eine „einseitige Community“, welches sich eindeutig widerspricht. Ein nicht außer Acht zu lassender Fakt ist auch die Bildung einer virtuellen Identität. Der Fragende und der Antwortende steigt bei Lycos iQ in seine eigens kreierte „virtuelle Haut“. Diese aufgebaute Identität, jederzeit wieder erkennbar durch den Usernamen und das Avatarbild, bekommt durch die gestellten Fragen sowie durch die abgegebenen Antworten einen eigenen Charakter. Das Ziel ist, ein geschätztes Community-Mitglied zu werden, welches sich unter anderem durch die konsequente Bewertung der Antworten seiner Frage auszeichnet [2]. Dies scheint maßgeblich die Begründung des Mangels an Fragen mit dem Status „Nicht bewertet“ zu sein.

3.3 Auswertung des Faktenfragenkatalogs

3.3.1 Auswertung der Faktenfragenkataloge

Der Auswertung lagen 100 untersuchte Faktenfragen zu Grunde. Auf diese 100 Faktenfragen wurde insgesamt 293 Mal geantwortet. Das entspricht einer durchschnittlichen Antwortenanzahl von 2,93. Aufgerundet wurde im Schnitt also auf jede Faktenfrage drei Mal geantwortet.

Doch wie viele der gegebenen Antworten waren für den Fragenden brauchbar? Von 293 Antworten waren insgesamt 235 Antworten (80,20 Prozent) nach Rechercheprüfung durch die Probanden richtig. Die Menge der als falsch bewerteten Antworten lag bei 57 (19,45 Prozent).

Insgesamt zwölf Antworten (4,1 Prozent), der Gesamtmenge wurden exakt durch denselben Wortlaut beantwortet. Die durchschnittliche Zeit, die der User, der als erstes auf eine Frage geantwortet hat, benötigte, liegt laut der Auswertung mit 208,23 Minuten sehr hoch. Dieses Ergebnis wurde von ein paar zeitlichen Ausreißern verzerrt. Durch den Median, welcher die oben genannten „Ausreißer“ ignoriert, ist zu errechnen, dass die durchschnittliche Antwortzeit, ohne die wenigen zeitlichen Ausreißer, bei drei Minuten liegt.

Es liegt im Interesse des Fragenden, eine möglichst umfangreiche Antwort zu bekommen. Entscheidend ist also, in welcher Form die Antworten gegeben werden. Die Antwortenden haben immer eine von vier Antwort-Formen gewählt. Zum einen gab es die Möglichkeit, ausschließlich mit einem Link zu antworten. Zum anderen konnte nur mittels eines Fließtextes geantwortet werden. Die dritte Möglichkeit war die Kombination aus Fließtext und Link und die vierte Form war das Beifügen eines Wikipedia-Links. Bei der folgenden Auflistung ist zu beachten, dass die gegebenen 293 Antworten 100 Prozent darstellen.

Von der ersten Form, dem ausschließlichen Beantworten durch einen Link, wurde neun Mal (3,07 Prozent) Gebrauch gemacht (Abb. 4.6). Die zweite Form einer Fließtext-Antwort wurde 270 Mal (92,15 Prozent) verwandt. Die Anzahl der Antworten, die durch eine Kombination aus Fließtext und Link formuliert wurden, liegt bei 67 Stück (22,87 Prozent). Die vierte Form der Beantwortung durch die Unterstützung eines Wikipedia-Links wurde 33 Mal (11,26 Prozent) in Anspruch genommen.

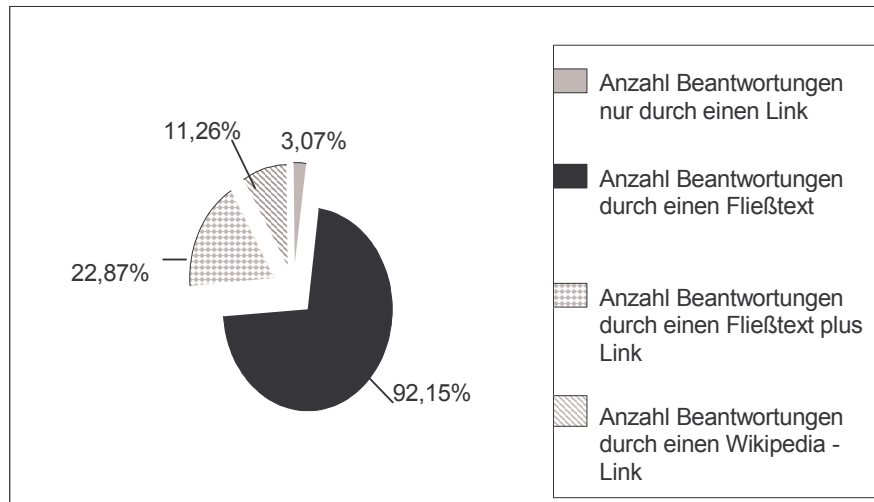


Abb. 4.6 Beantwortung allgemeiner Faktenfragen (n=293)

Die genaue Betrachtung dieser Werte gibt deutlich preis, dass die Antwort mittels eines Fließtextes die häufigste Form der Beantwortung war. Dies lässt auf zwei Begründungen schließen. Zum einen ist es möglich, dass die Antwortenden im Bereich Ihres Themengebietes geantwortet haben und somit die Antwort ihrem Allgemeinwissen entnommen haben. Zum anderen ist es möglich, dass die Antwort einer Quelle entnommen wurde, welche aber nicht genannt wurde.

Die zweithäufigste Form der Beantwortung ist die Kombination Fließtext und Link. Dies lässt darauf schließen, dass die Antwortenden ihrer Antwort entweder eine Quellenangabe hinzugefügt haben oder mehr Tiefe verleihen wollten, indem sie einen Link zu einer Seite hinzufügt haben, die weitere Inhalte zu diesem Thema aufführt. So kann der Fragende selber entscheiden, ob die angegebenen Informationen in dem Fließtext für ihn ausreichend sind oder ob er sich über den angegebenen Link weiter informieren möchte. Beide Erklärungen zeigen, dass es sinnvoll ist, diese Variante der Antwort zu wählen. Diejenigen, die diese Form der Beantwortung gewählt haben, hatten wahrscheinlich das Ziel, eine möglichst genaue Antwort zu geben.

Die dritthäufigste Form der Beantwortung ist die Antwort mittels eines Wikipedia- Links. Diese Antwort schließt allerdings alle Antworten mit ein, die eventuell auch in Kombination eines Fließtextes gegeben worden sind oder auch die Antworten, welche ausschließlich durch einen Wikipedia-Link beantwortet worden sind. Die Untersuchung dieser Frage sollte überprüfen, wie oft die Antwortenden auf Wikipedia verweisen und ob somit eine Vielzahl der gegebenen Antworten ohne den Umweg über Lycos iQ auskommen würden. Da sich diese Zahl aber laut der durchgeführten Stichprobe, nur knapp über 10 Prozent bewegt, ist diese Vermutung nicht zutreffend.

Die am wenigsten gewählte Form der Beantwortung ist die der ausschließlichen Linkangabe. Hier besteht die Vermutung, dass die Begründung dafür wieder mit dem „Geist“ der Community zu erklären ist. Diese Form enthält die geringste Chance auf eine der drei positiven Bewertungen anderer User oder des Fragenden. Hier kann auch keine persönliche Note durch Wortwitz oder Emoticon wie bspw. Smileys eingebaut

und somit auch nicht der virtuelle Charakter gegenüber anderen Community-Mitgliedern positiv geformt werden. Der Antwortende, der sich für diese Form entscheidet, will möglicherweise nur schnell und unpersönlich sein Wissen weiter geben.

3.3.2 Linkangabe in den Antworten

Ganz allgemein wurden 109 Antworten durch einen Link gegeben. Diese Links sind von unterschiedlicher Qualität und wurden zur besseren Übersichtlichkeit in die drei Kategorien „Top Quellen“, „Offizielle Quellen“ und „Inoffizielle Quellen“ eingeordnet.

Die Einstufung in diese Kategorien soll vor allem die Qualität der Quellen deutlich machen. Von den angegebenen 109 Antworten der Untersuchung enthielten 76 Antworten einen Link. Allerdings wurden von den Probanden nur 57 Links kategorisiert. Es stammen sieben Links (12,28 Prozent) aus der Kategorie eins (Top Quellen), 23 Links (40,35 Prozent) aus der Kategorie zwei (Offizielle Quellen) und 27 Links (47,37 Prozent) aus der Kategorie drei (Inoffizielle Quellen). Hier sieht man deutlich, dass die Angabe der inoffiziellen Links am häufigsten gewählt wurde. Die zweithäufigst gewählte war die Kategorie zwei und die am wenigsten gewählte war die erste Kategorie. Dies lässt den Schluss zu, dass die angegebenen Links am häufigsten zu wenig zuverlässigen Quellen führen. Da die verlässlichen und die unverlässlichen Quellen zu ungefähr gleichen Teilen angegeben wurden, ist die Angabe eines Links demnach weder ein Garant für eine richtige Antwort, noch ein Indiz für eine qualitativ hochwertige Antwort, entnommen einer verlässlichen Quelle. Allerdings wird nicht ausgeschlossen, dass der Antwortende seine Antwort mit der Angabe einer Quelle bestücken wollte, um eine formal korrekte Antwort zu geben und auch nicht, dass der Antwortende seine Antwort mit dem angegebenen Link unterstreichen wollte. Hier ist zu beachten, dass dem Antwortenden die Verlässlichkeit der Quelle eventuell nicht bewusst ist und er somit unwissend eine nicht verlässliche Quelle angegeben hat.

Schlussendlich kann man sagen, dass etwas weniger als die Hälfte der Gesamtanzahl angegebener Links des Tests aus unverlässlichen Quellen stammt. Die restlichen Links verteilen sich auf 12,28 Prozent der Links aus Top-Quellen und 40,35 Prozent der Links aus offiziellen Quellen. Diese Links sind Quellen mit hoher bis zufrieden stellender Verlässlichkeit. Addiert ergeben diese Quellen mit einer Prozentzahl von 52,63 Prozent sogar eine größere Summe als die der nicht seriösen Quellen mit 47,37 Prozent. Diese Auflistung zeigt deutlich, dass die Angabe von verlässlichen Links häufiger geschieht als die von Unverlässlichen.

3.3.3 Bewertung abgegebener Antworten

Um als Fragender dem Antwortenden und auch den anderen Community-Mitgliedern deutlich zu machen, für wie nützlich eine abgegebene Antwort empfunden wird, gibt es verschiedene Bewertungsmöglichkeiten. Diese Bewertungen können mehrfach innerhalb aller abgegebenen Antworten einer Frage vergeben werden. Der Fragende kann hier zwischen Top-Antwort, hilfreicher Antwort oder nicht hilfreicher Antwort wählen. Ebenso steht es dem Fragenden offen die Antwort nicht zu bewerten (Status nicht bewertete Antworten).

Auch die anderen Community-Mitglieder haben die Chance, die abgegebenen Antworten einer Frage mit „gut“ zu bewerten.

Die 293 abgegebenen Antworten wurden insgesamt 354 Mal bewertet. Die bewertete Zahl ist deshalb höher, weil es möglich ist, eine Antwort durch den Fragenden und ebenso von anderen Community-Mitgliedern bewerten zu lassen. Von den untersuchten 293 Antworten wurden 95 (32,42 Prozent) als Top-Antwort bewertet. Insgesamt 158 von 293 Antworten (53,92 Prozent) wurden als hilfreich bewertet. Die Anzahl der richtigen Antworten, die weder als Top-Antwort noch als hilfreiche Antwort bewertet wurden und die auch nicht von anderen Community-Mitgliedern bewertet wurden, beträgt 57 (19,45 Prozent). Somit wurden 44 Antworten (15,02 Prozent) von anderen Community-Mitgliedern als gut bewertet.

Es wird davon ausgegangen, dass die von anderen Usern als gut bewerteten Antworten auch gleichzeitig richtig sind. Die Auswertung zeigt aber, dass dies nicht der Regelfall ist. Von 44 gut bewerteten Antworten sind 35 auch richtig und neun sind falsch. Dies zeigt, dass eine solche Bewertung mit der Richtigkeit einer Antwort nicht gleich zu setzen ist.

3.3.4 Verteilung der Antwortgeber

Um einen Eindruck von den Antwortenden zu erhalten wird im Folgenden die gesamte Rangverteilung näher betrachtet (vgl. Abb. 4.7).

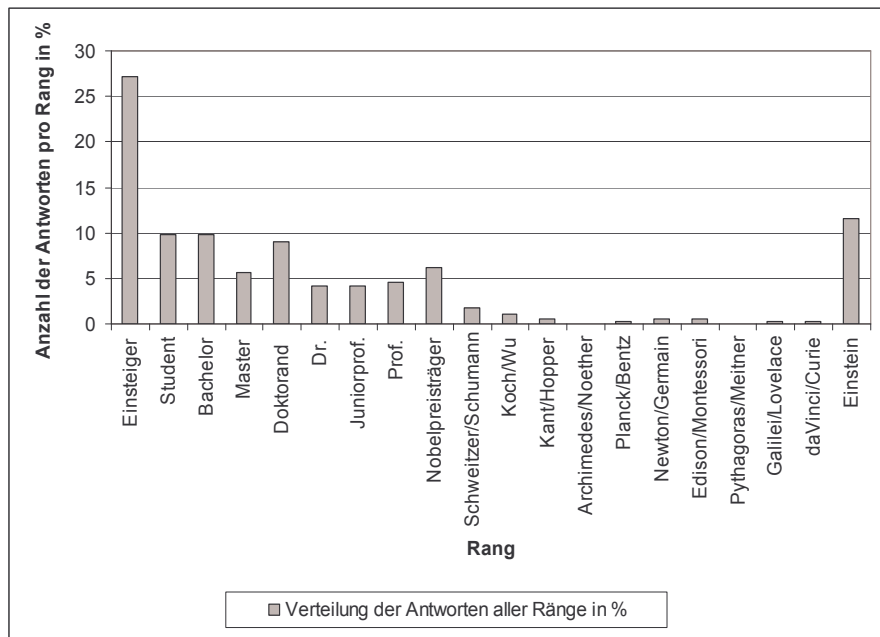


Abb. 4.7 Verteilung der Antworten pro Rang (n=354)

(Faktenfragen)

Die Verteilung der 20 Ränge bei den untersuchten Faktenfragen zeigt, dass der Rang Einsteiger mit circa 27 Prozent der Antworten im Bereich der Faktenfragen der am häufigsten vertretene Rang ist. Die Abbildung weist einen starken Abfall der Vertreter der übrigen Ränge an. Hier können folgende Vermutungen angestellt werden:

Es existieren mehr User mit dem Rang Einsteiger, da es sich hier um User handelt, die sich erst vor kurzem bei Lycos iQ angemeldet haben und noch nicht die Möglichkeit hatten, einen oder mehrere Ränge aufzusteigen. Dies würde bedeuten, dass im Moment ein starker Aufschwung der Neuanmeldungen bei Lycos iQ stattfindet.

Es könnte aber auch daraufhin deuten, dass viele User, nachdem sie durch den Einstieg bei Lycos iQ den Rang Einsteiger erworben haben, kein Interesse mehr an dem Erreichen eines höheren Ranges bei Lycos iQ haben. Hierfür würde z.B. auch die geringe Anzahl an Usern in den übrigen Rängen sprechen.

Die geringe Menge der anderen Ränge steht dem hohen Anteil des Ranges Einstein gegenüber. Er stellt den zweithäufigsten Rang mit fast 12 Prozent der Antworten bei der Betrachtung der gesamten Rangverteilung dar. Dies zeigt, dass sich eine gewisse Anzahl an Mitgliedern diesen höchsten Rang erarbeitet hat. Diese Mitglieder nehmen aktiv an dem Community-Leben bei Lycos iQ teil.

Es liegt also die Vermutung nahe, dass es auf der einen Seite bei Lycos iQ Mitglieder gibt, welche sich stark engagieren, also viele Antworten geben und daran interessiert sind, ihren Expertenstatus zu verbessern. Auf der anderen Seite gibt es den (Neu-)Einsteiger bei Lycos iQ, welcher wenig Interesse an dem Frage-Antwort-Dienst zeigt und nach kurzer Zeit nicht mehr bei Lycos iQ aktiv mitwirkt. Weiterhin gibt es auch diejenigen, der doch an einem aktiven Mitwirken in der Community und der Verbesserung seiner Ranges interessiert ist und daher schnell den Rang Einsteiger verlässt und sich in Folge seines Antwortgebens bald auf einem der höheren Ränge wieder findet.

3.3.5 Verteilung der Ränge der Antwortenden bei der am schnellsten abgegebenen Antwort

Die jeweils am schnellsten abgegebenen Antworten der 100 untersuchten Fragen wurden genau überprüft. Unter anderem wurde notiert, welchen Rang der Antwortende hat, welcher die jeweils schnellste Antwort abgegeben hat. Es ist zu beachten, dass von den 100 untersuchten Usern zwei keinen Rang innehatten oder dieser von den Probanden nicht angegeben wurde.

Bei der Auswertung war deutlich zu erkennen, dass von den 20 möglichen Rängen die User mit den Rängen Einsteiger, Einstein und Student am häufigsten die schnellste Antwort abgegeben haben.

Um die folgenden Grafiken (Abb. 4.8 – Abb. 4.15) verständlich und lesbar zu machen, erfolgt nun eine genaue Erklärung derselbigen anhand des vorliegenden Beispiels. Die Angaben in der Grafik werden in Prozent gegeben, dabei zeigt die X-Achse den Rang, welchen der Antwortgeber innehat, und die Y-Achse verzeichnet die Anzahl an Antworten pro Rang in Prozent.

Die Balken in der Grafik stellen die Verteilung aller untersuchten Antworten pro Rang dar. So kann man in Abb. 4.8 ablesen, dass der Rang Einsteiger circa 27 Prozent

von allen untersuchten Antworten gegeben hat, die nachfolgenden Ränge Student und Bachelor hingegen nur jeweils circa 10 Prozent aller Antworten.

Vor diesem Hintergrund erfolgt im Beispiel die Verteilung der schnellsten Antworten auf die Ränge in Form einer Linie. Auch hier wurden die Angaben in Prozent gemacht und errechnen sich an der Anzahl aller untersuchten Antworten. Abb. 4.8 zeigt hier zum Beispiel, dass die Mitglieder mit dem Rang Einsteiger am häufigsten die schnellste Antwort liefern (mit circa 7 Prozent). Da sich aber im Vergleich zu der gesamten Rangverteilung zeigt, dass dieser Rang in einer viel größeren Masse vorhanden ist als z.B. der Rang Nobelpreisträger/in, ist dieser folglich auch bei dem Anteil an schnellsten Antwort häufiger vertreten als andere Ränge. Es kann also kein bestimmter Rang als schnellster Antwortgeber ausgezeichnet werden.

Diese Erläuterungen können auf alle folgenden Grafiken, welche sich mit der Bewertung der Antworten beschäftigen, bezogen werden.

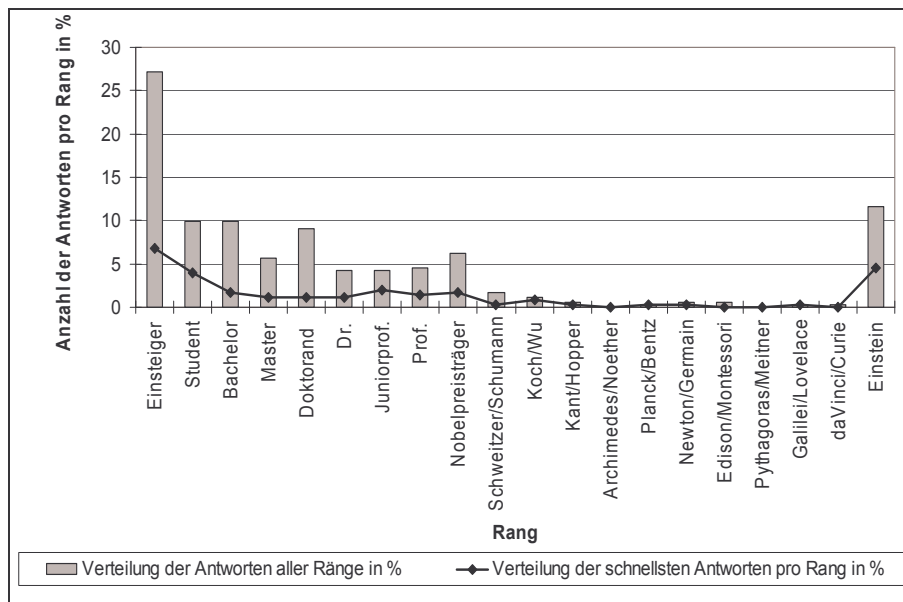


Abb. 4.8 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung schnellste Antwort (n=98) (Faktenfragen)

3.3.6 Untersuchung der richtigen Antworten, die weder als Top, hilfreich oder gut bewertet wurden

Die Antworten, die weder als Top, hilfreich oder gut bewertet wurden, haben eine Gesamtanzahl von 57. Die Antwortenden dieser 57 Antworten konnten 20 mögliche Ränge innehaben.

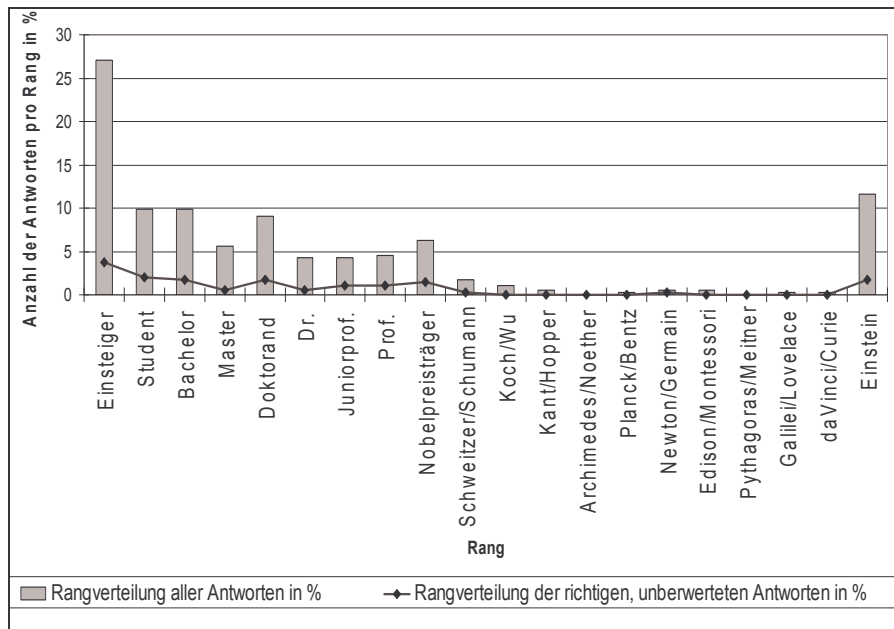


Abb. 4.9 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung richtige, unbewertete Antworten (n=57) (Faktenfragen)

Die unbewerteten Antworten, die durch Recherche der Probanden als richtig eingestuft wurden, konnten nun mit der gesamten Rangverteilung verglichen werden. Dadurch deuten sich folgende Ausprägungen an:

Die Ränge, die eine höhere Anzahl haben, wie z.B. Einsteiger, sind auch am häufigsten unter den richtigen, unbewerteten Antworten vertreten (4 Prozent gemessen an allen untersuchten Antworten). Lediglich der Rang Einstein (2 Prozent) weicht von der Darstellung der gesamten Rangverteilung ab.

Dies könnte heißen, dass Antworten von Mitgliedern mit dem Rang Einstein eher unter den bewerteten Antworten zu finden sind. Ein Grund dafür könnte sein, dass die Antworten des Ranges Einstein richtig und eventuell so gut sind, dass die Fragesteller dies mit entsprechender Bewertung berücksichtigen wollen.

Jeder User von Lycos iQ hat die Möglichkeit Themengebiete auszuwählen. Dies soll anderen Usern zeigen, dass er ein besonderes Interesse an diesem Thema hat oder besondere Fähigkeiten besitzt, welche ihn prädestinieren, in diesem Bereich zu antworten. Von 57 abgegebenen Antworten wurden 18 (31,58 Prozent) von Usern beantwortet, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. 32 Antworten (56,14 Prozent) wurden von Usern gegeben, die das jeweilige Themengebiet nicht als das Eigene angegeben haben. Bei sieben Antworten (12,28 Prozent) konnte keine derartige Auswertung stattfinden, da die User mitsamt ihrer Profile zur Zeit der Auswertung bereits deaktiviert waren; ihre Antworten sind aber zur Ansicht noch aufrufbar.

Die genaue Untersuchung der Antworten, die weder als Top, hilfreich oder gut bewertet wurden, soll aufzeigen, in welcher Form geantwortet wurde.

Von den 57 untersuchten Antworten wurden zwei (3,51 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 44 Mal (77,19 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde 14 Mal (24,56 Prozent) genutzt. Davon waren acht (14,04 Prozent) der angegebenen Links Wikipedia-Links. Hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Allerdings ergaben die Antworten, die ausschließlich durch einen Link gegeben wurden, die, welche einen Fließtext plus Link enthielten, und die Wikipedia-Link-Antworten insgesamt eine Anzahl von 24 Antworten, welche einen Link enthielten. Dies zeigt, dass die Vergabe von Links in den Antworten nicht die Ausnahme ist, sondern nahezu in der Hälfte der Antworten verwendet wurde. Die regelmäßige Wahl dieser Form lässt darauf schließen, dass sich diese Art der Antwort bewährt hat.

Abzüglich der Wikipedia-Links sind es 16 Links, welche noch einmal separat betrachtet und extra kategorisiert werden. Zwei dieser Links können der ersten Kategorie zugeordnet werden, vier der zweiten Kategorie und zehn der dritten. Diese Kategorisierung zeigt die Tendenz, dass hier die häufigste Quelle die „Inoffiziellen Quellen“ sind.

3.3.7 Untersuchung der Top Antworten

Die Top Antworten haben eine Gesamtanzahl von 95. Die Antwortenden dieser 95 Antworten konnten 20 verschiedene Ränge innehaben.

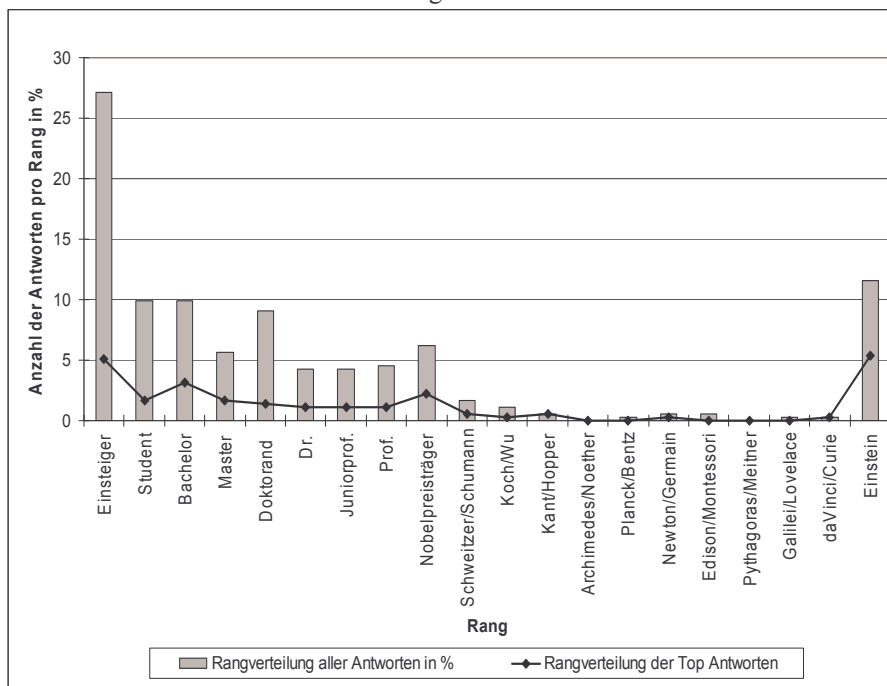


Abb. 4.10 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung Top bewertete Antworten (n=95) (Faktenfragen)

Bei der Rangverteilung der Top bewerteten Antworten zeigt sich eine ähnliche Ausprägung der Ränge, wie bei der Gesamtaufteilung aller untersuchten Antworten pro Rang bei Lycos iQ (vgl. Abb. 4.10, hier als Balken). Die Linie der Rangverteilung der Top bewerteten Antworten zeigt deutlich, dass hier die meisten Antworten von den Rängen Einstein (5,37 Prozent) und Einsteiger (5,08 Prozent) gegeben wurden. Die anderen Ränge sind in geringerer Anzahl oder gar nicht vertreten. Die meisten Top Antworten werden folglich von den Rängen Einstein und Einsteiger gegeben, bei einem Vergleich mit der Verteilung aller Antworten wird darüber hinaus deutlich, dass es sich bei fast der Hälfte aller von dem Rang Einstein gegebenen Antworten um Top Antworten handelt. Man kann hier auf Grund der niedrigen Fallzahlen (nur 41 aller Antworten wurden von dem Rang Einstein gegeben) lediglich von einer leichten Korrelation zwischen dem hohen Rang Einstein und Top Antworten sprechen.

Außerdem fehlen bei den dem Rang Einstein vorgelagerten Rängen (z.B. Da Vinci/ Curie) ebenfalls Fallzahlen, welche die Vermutung einer Korrelation unterstreichen würden. Denn bei einem hohen Anteil von Top Antworten bei dem Rang Einstein wäre es auch sehr wahrscheinlich, dass die Ränge, welche kurz vor dem Rang Einstein liegen, ebenfalls eine überdurchschnittlich hohe Menge an Top Antworten liefern.

Insgesamt wurden 95 Antworten von den Fragenden als Top bewertet. Davon waren 88 (92,63 Prozent) richtig und sieben falsch (7,37 Prozent) (siehe zur Definition von richtig und falsch Punkt „Allgemeine Fragen“ in Kap. 2.2.2).

Es gab in dieser Kategorie insgesamt 34 Antworten (35,79 Prozent), die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 57 Antworten (60 Prozent) wurde von Usern gegeben, die das jeweilige Themengebiet nicht als das Eigene angegeben haben. Bei vier Antworten (4,21 Prozent) konnte keine derartige Auswertung stattfinden, da die User mitsamt ihrer Profile zur Zeit der Auswertung bereits deaktiviert waren.

Von den 95 untersuchten Antworten wurden sechs (6,32 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 84 Mal (88,42 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde 34 Mal (35,79 Prozent) benutzt. Davon waren zwölf (12,63 Prozent) der angegebenen Links Wikipedia-Links. Auch hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden, die, die einen Fließtext plus Link enthielten, und die Wikipedia-Link Antworten ergeben hier insgesamt eine Anzahl von 52 Antworten, welche einen Link enthielten. Dies zeigt, dass auch hier die Vergabe von Links in den Antworten nicht die Ausnahme ist.

Abzüglich der Wikipedia-Links sind es 40 Links, welche noch einmal separat betrachtet und extra kategorisiert werden. Allerdings ist hier zu beachten, dass von den Probanden nur 32 Links einkategorisiert wurden. Es wurden sechs dieser Links aus der ersten Kategorie entnommen, 14 wurden der zweiten Kategorie und zwölf der dritten entnommen. Diese Kategorisierung zeigt die Tendenz, dass hier die häufigste Quelle der Links die „Offiziellen Quellen“ sind.

3.3.8 Zusammenhang von Rang und richtiger Antwort bei den Top bewerteten Antworten

Die Überprüfung des Zusammenhangs von der Richtigkeit einer Antwort und dem Rang des Antwortenden war ein Hauptbestandteil dieser Untersuchung. Hierfür wurden die drei Bewertungsmöglichkeiten Top-Antwort, hilfreiche Antwort und gut bewertete Antwort separat geprüft.

Zunächst wurden die verschiedenen Antwortvarianten der als Top-Antwort bewerteten Antworten verglichen. Von 84 Antworten, welche mit einem Fließtext beantwortet wurden, waren 77 Antworten (91,67 Prozent) richtig. Alle sechs Antworten, welche durch einen Link beantwortet wurden, waren richtig. Die 34 Antworten, welche durch einen Fließtext und durch einen Link beantwortet wurden, waren 32 Mal (94,12 Prozent) richtig und die zwölf durch Wikipedia-Links beantworteten Antworten waren alle richtig.

Es wurden bei den Top-Antworten insgesamt zu 93 Prozent richtige Antworten gegeben. Da die Abweichung von 7 Prozent sehr gering ist, kann man darauf schließen, dass eine als Top bewertete Antwort mit großer Wahrscheinlichkeit auch richtig ist.

Die Überprüfung der Beziehung zwischen Top-Antwort und dem Rang des Antwortenden soll Aufschluss über einen möglichen Zusammenhang zwischen einem hohen Rang und einer richtigen Antwort geben (vgl. Abb. 4.11).

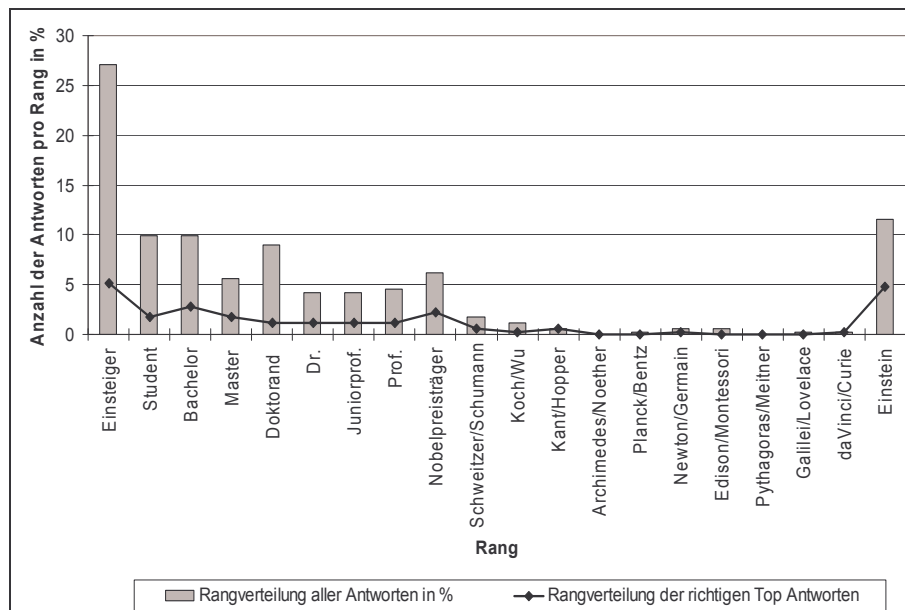


Abb. 4.11 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung der richtigen, Top bewerteten Antworten (n=88) (Faktenfragen)

Die Grafik zeigt, dass kein deutlicher Zusammenhang zwischen dem Rang eines Antwortenden und einer richtigen Antwort erkennbar ist. Vielmehr kann man sagen, dass die Rangverteilung hier nur gering von der gesamten Rangverteilung abweicht.

Bei der gesamten Rangverteilung war der Einsteiger an erster Stelle. Bei der Rangverteilung der richtigen Top Antworten zeigte sich ebenso der Einsteiger auf diesem Platz, doch der Rang Einstein folgte mit nur einer Antwort weniger. Dies zeigt, dass eventuell eine Korrelation zwischen richtiger Antwort und dem Rang Einstein festzuhalten ist. Mit Sicherheit kann dies allerdings nicht behauptet werden, da dazu die Fallzahlen zu gering sind (siehe dazu auch Kap. 3.4.5).

Von 34 Top-Antworten, welche der Antwortende in seinem eigenen Themengebiet beantwortet hat, fielen 31 Antworten (91,18 Prozent) richtig aus. Allerdings haben die 54 Mitglieder, welche nicht in ihrem Themengebiet geantwortet haben, auch 51 richtige Antworten gegeben, was einen Prozentsatz von ca. 94 Prozent richtigen Antworten darstellt.

Es kann also kein direkter Zusammenhang zwischen den eigens erwählten Themengebieten und der Richtigkeit einer Top-Antwort nachgewiesen werden. Beide Antwortarten sind, zu ungefähr gleich hohen Anteilen, mit richtigen Antworten bestückt worden.

Es stellt sich außerdem die Frage, ob eine vom Fragesteller als Top bewertete Antwort ebenso von den anderen Community-Mitgliedern als eine brauchbare Antwort angesehen wird. Dies wird bei der Untersuchung der Korrelation von Top-Antwort und gut bewerteter Antwort in Hinsicht auf die Richtigkeit deutlich. Von insgesamt 95 Top Antworten waren 88 richtig. Von diesen 88 wurden 33 (68,18 Prozent) jeweils einmal von anderen Usern als gut bewertet. Insgesamt sieben Antworten (15,91 Prozent) wurden jeweils zweimal und fünf (11,36 Prozent) wurden jeweils dreimal als gut bewertet. Eine viermalige gute Bewertung durch andere User erhielten zwei Top-Antworten (4,55 Prozent).

Insgesamt wurden 44 Antworten (50 Prozent) von dem Fragenden als Top und gleichzeitig von anderen Usern als gut bewertet. Es bleibt festzuhalten, dass hierbei nicht von einer Korrelation beider Bewertungsmöglichkeiten gesprochen werden kann.

3.3.9 Untersuchung der hilfreichen Antworten

Die hilfreichen Antworten haben eine Gesamtanzahl von 158. Die Antwortenden dieser 158 Antworten konnten 20 mögliche Ränge innehaben.

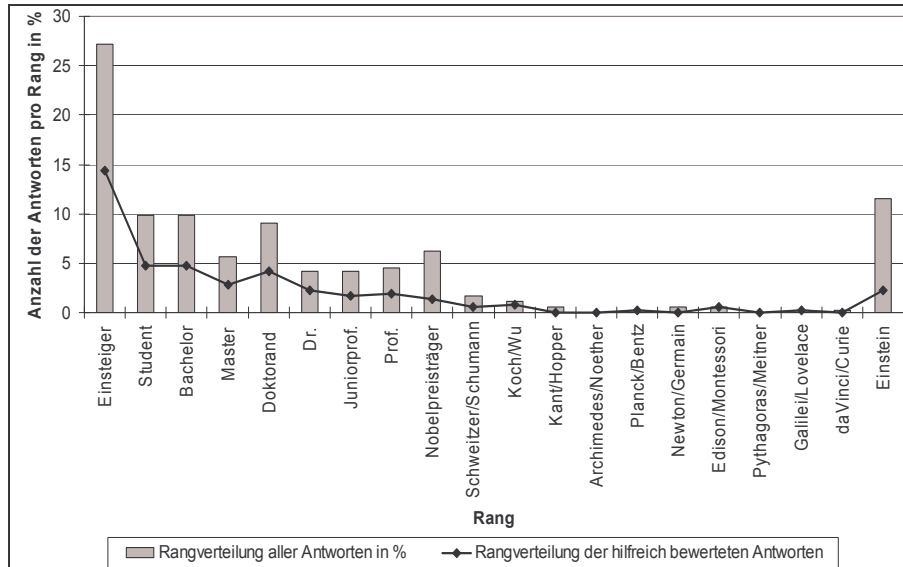


Abb. 4.12 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung der hilfreich bewerteten Antworten (n=158) (Faktenfragen)

Vergleicht man die Rangverteilung der hilfreichen Antworten (Linie) mit der gesamten Rangverteilung aller Antworten (Balken) bei Lycos iQ, fällt auch hier wieder auf, dass die Ränge, die die meisten hilfreichen Antworten gegeben haben, auch bei der Betrachtung der gesamten Rangverteilung als die am häufigsten vertretenen aufgefallen sind. So hat der Rang Einsteiger mit einem Anteil von circa 14 Prozent an allen untersuchten Antworten die meisten hilfreichen Antworten gegeben. Bei dem Rang Einsteiger handelt es sich aber ebenso um den am meisten vertretenen Rang.

Anders als bei der gesamten Rangverteilung fällt hier der Rang Einstein durch eine geringe Anzahl (52,26 Prozent) an hilfreichen Antworten auf. Betrachtet man hingegen die gesamte Aufteilung der Ränge, ist der Einstein an zweiter Stelle. Dies deutet darauf hin, dass Mitglieder, welche den Rang Einstein innehaben, wenige hilfreiche Antworten, sondern vielmehr bessere Antworten (Top Antworten) geben. Es könnte aber auch zeigen, dass Mitglieder mit dem Rang Einstein eher besser bewertet werden, da man ihre Antworten aufgrund ihres Ranges potentiell als Top Antworten einstuft. Dies spricht für die bereits angedeutete eventuelle Korrelation von richtiger Antwort und dem Rang Einstein.

Insgesamt wurden 158 Antworten von den Fragenden als hilfreich bewertet. Davon waren 117 (74,05 Prozent) richtig und 36 (22,78 Prozent) falsch. Es ist zu beachten, dass fünf Antworten (3,16 Prozent) von den Probanden nicht bearbeitet worden sind.

Es gab in dieser Kategorie insgesamt 43 (27,22 Prozent) Antworten, die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 100 Antworten (63,29 Prozent) wurde von Usern gegeben, die das jeweilige Themengebiet nicht als das eigene angegeben haben. Bei zehn Antworten (6,33 Prozent) konnte keine derartige Auswertung

stattfinden, da die User mitsamt ihrer Profile zur Zeit der Auswertung bereits deaktiviert waren; ihre Antworten sind aber zur Ansicht noch aufrufbar. Außerdem ist auch hier zu beachten, dass fünf Antworten (3,16 Prozent) von den Probanden nicht bearbeitet worden sind.

Von den 158 untersuchten Antworten wurden drei (1,96 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 137 Mal (89,54 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde 27 Mal (17,65 Prozent) benutzt. Davon waren 13 (8,5 Prozent) der angegebenen Links Wikipedia-Links. Auch hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden, jene die einen Fließtext plus Link enthielten und die Wikipedia-Link-Antworten ergeben hier insgesamt eine Anzahl von 43 Antworten, welche einen Link enthielten.

Abzüglich der Wikipedia-Links sind es 30 Links, welche noch einmal separat betrachtet und kategorisiert werden. Allerdings ist hier zu beachten, dass von den Probanden nur 21 Links kategorisiert wurden. Es wurde ein Link aus der ersten Kategorie entnommen, elf Links wurden der zweiten Kategorie und neun Links der dritten Kategorie entnommen. Diese Kategorisierung zeigt die Tendenz, dass die häufigste Quelle der Links die „Offiziellen Quellen“ sind.

3.3.10 Zusammenhang von Rang und richtiger Antwort bei den als hilfreich bewerteten Antworten

Auch an dieser Stelle der Untersuchung wurde erneut der Zusammenhang von der Richtigkeit einer Antwort und dem Rang des Antwortenden überprüft. Zunächst wurden die verschiedenen Antwortvarianten der als hilfreich bewerteten Antworten verglichen. Von 137 Antworten, welche mit einem Fließtext beantwortet wurden, waren 102 (74,45 Prozent) richtig.

Von drei Antworten, welche durch einen Link beantwortet wurden, waren alle richtig. Die 27 Antworten, welche durch einen Fließtext und durch einen Link beantwortet wurden, und die 13 durch Wikipedia-Links beantworteten Antworten waren ebenfalls alle richtig.

Es wurden bei den hilfreichen Antworten insgesamt zu 74 Prozent richtige Antworten gegeben. Die Abweichung von 26 Prozent lässt lediglich die Wahrscheinlichkeit zu, dass die Richtigkeit einer Antwort bei den als hilfreich bewerteten Antworten höher ist, als die Wahrscheinlichkeit, dass sie falsch ist.

Die Überprüfung der Beziehung zwischen einer als hilfreich bewerteten Antwort und dem Rang des Antwortenden soll Aufschluss über einen möglichen Zusammenhang zwischen einem hohen Rang und einer richtigen Antwort geben (vgl. Abb. 4.13).

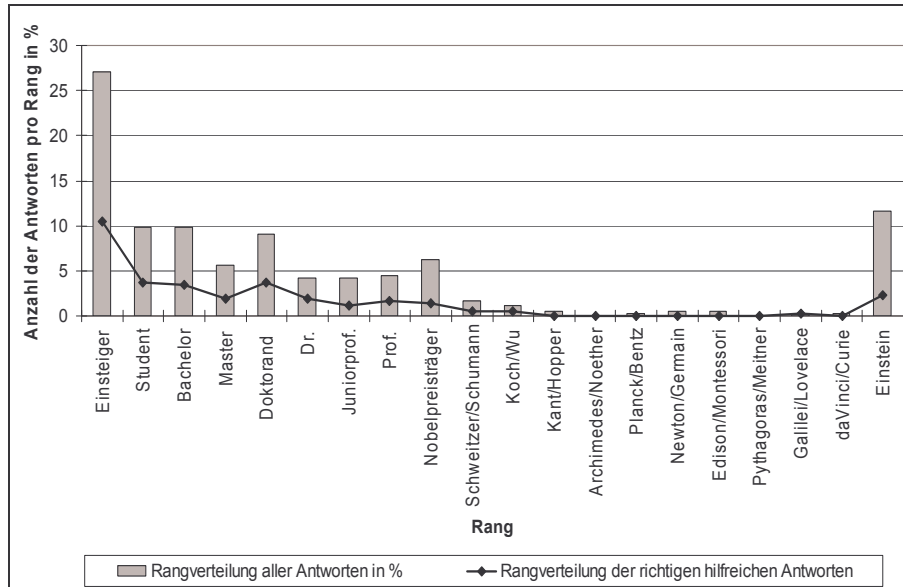


Abb. 4.13 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung der richtigen, hilfreich bewerteten Antworten (n=117) (Faktenfragen)

Die Grafik zeigt, dass in diesem Fall kein Zusammenhang zwischen dem Rang eines Antwortenden und einer richtigen hilfreichen Antwort erkennbar ist. Vielmehr wird hier durch die Linie die Rangverteilung aller Antworten (Balken) erneut abgebildet, dass der Einsteiger auch hier an erster Stelle der Rangverteilung liegt, gefolgt von den Rängen Doktorand und Student. Eine Besonderheit bei den richtigen hilfreichen Antworten ist, dass sich der Rang Einstein hier nicht, wie sonst üblich, an zweiter Position befindet. Vermutungen diesbezüglich lassen sich in dem vorangegangenen Kapitel finden.

Die Deutung dieser Ergebnisse lässt den Schluss zu, dass diejenigen Ränge, welche am meisten vorhanden sind, auch am meisten richtige Antworten produzieren. Hierbei stellt der Rang Einstein, wie bereits beschrieben, eine Ausnahme dar. Dies könnte ein Hinweis auf eine Korrelation zwischen Einstein und hohem Rang sein. Von den 43 als hilfreich bewerteten Antworten, welche der Antwortende in seinem eigenen Themengebiet beantwortet hat, fallen 33 (76,74 Prozent) Antworten richtig aus.

Zu klären bleibt jedoch, ob eine vom Fragesteller als hilfreich bewertete Antwort ebenso von den anderen Community-Mitgliedern als eine brauchbare Antwort angesehen wird. Dies wird bei der Untersuchung der Korrelation von hilfreicher Antwort und gut bewerteter Antwort in Hinsicht auf die Richtigkeit deutlich. Von insgesamt 158 hilfreichen Antworten waren 117 richtig. Von diesen 117 wurden insgesamt 28 Antworten (23,93 Prozent) von dem Fragenden als hilfreich und gleichzeitig von anderen Usern als gut bewertet.

Von diesen 28 Antworten wurden 21 Antworten (75 Prozent) jeweils einmal von anderen Usern als gut bewertet. Insgesamt fünf Antworten (17,86 Prozent) wurden jeweils zweimal und zwei Antworten (7,14 Prozent) wurden jeweils dreimal als gut

bewertet. Eine viermalige gute Bewertung durch andere User wurde in diesem Teil der Untersuchung nicht vergeben.

Es bleibt festzuhalten, dass hierbei nicht von einer Korrelation beider Bewertungsmöglichkeiten gesprochen werden kann.

3.3.11 Untersuchung der Antworten, die von anderen Usern als gut bewertet wurden

Die als gut bewerteten Antworten haben eine Gesamtanzahl von 44. Bei der Betrachtung der von anderen Usern als gut bewerteten Antworten (Linie) im Vergleich mit der Verteilung aller untersuchten Antworten auf die Ränge (Balken) zeigt sich wiederholt eine ähnliche Mengenverteilung der Ränge, wie bereits in den vorangegangenen Kapiteln beschrieben (vgl. Abb. 4.14). Auch hier sind die Mitglieder mit dem Rang Einsteiger am häufigsten vorhanden (3,95 Prozent), gefolgt von dem Rang Einstein (2,26 Prozent). Dies deutet also daraufhin, dass die Ränge mit den höchsten Mitgliederzahlen auch die meisten Antworten geben.

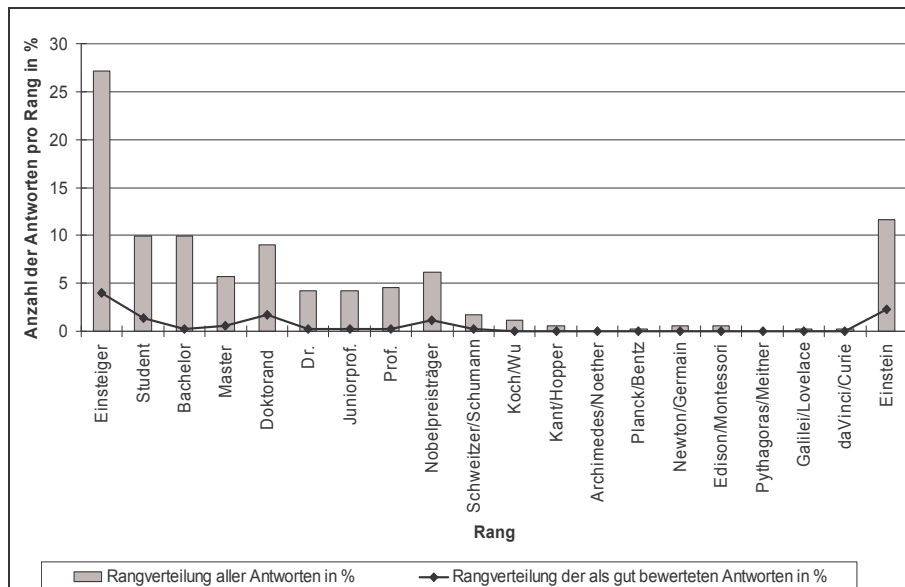


Abb. 4.14 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung der gut bewerteten Antworten (n=44) (Faktenfragen)

Von den 44 Antworten, die von anderen Usern als gut bewertet wurden, waren 35 (79,55 Prozent) richtig und neun (22,78 Prozent) falsch.

Es gab in dieser Kategorie insgesamt 16 Antworten (36,36 Prozent), die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 27 Antworten (61,36 Prozent) wurden von Usern gegeben, die das jeweilige Themengebiet nicht als das eigene angegeben haben. Bei einer Antwort (2,27 Prozent) konnte keine derartige Auswertung stattfinden, da der User mitsamt seinem Profil zur Zeit der Auswertung bereits deaktiviert war.

Von den 44 untersuchten Antworten wurden drei (6,82 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 38 Mal (86,36 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde fünf Mal (11,36 Prozent) benutzt. Davon war einer (2,27 Prozent) der angegebenen Links ein Wikipedia-Link. Hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden, jene die einen Fließtext plus Link enthielten und die Wikipedia-Link-Antworten ergeben hier insgesamt eine Anzahl von neun Antworten, welche einen Link enthielten.

Abzüglich des Wikipedia-Links sind es acht Links, welche noch einmal separat betrachtet und extra einkategorisiert wurden. Es wurden drei Links aus der ersten Kategorie entnommen, vier Links wurden der zweiten Kategorie und ein Link der dritten entnommen. Diese Kategorisierung zeigt die Tendenz, dass auch hier die häufigste Quelle der Links die „Offiziellen Quellen“ sind.

3.3.12 Zusammenhang von Rang und richtiger Antwort bei den als gut bewerteten Antworten

Auch an dieser Stelle der Untersuchung wurde der Zusammenhang von der Richtigkeit einer Antwort und dem Rang des Antwortenden überprüft. Zunächst wurden die verschiedenen Antwortvarianten der als gut bewerteten Antworten verglichen. Von 38 Antworten, welche mit einem Fließtext beantwortet wurden, waren 29 Antworten (76,32 Prozent) richtig. Von drei Antworten, welche durch einen Link beantwortet wurden, waren alle richtig. Die fünf Antworten, welche durch einen Fließtext und durch einen Link beantwortet wurden, waren vier Mal (80, Prozent) richtig. Die eine Antwort, die durch einen Wikipedia-Link beantwortet wurde, war ebenso richtig.

Da die Zahlen aus der Untersuchung gering sind, kann man lediglich folgende Tendenz feststellen: Die etwas größeren Abweichungen bei den richtigen Antworten in Fließtextform lassen darauf schließen, dass eine gut bewertete Antwort nicht zwingend ein Garant für die Richtigkeit der Antwort ist, wenn sie in Fließtext-Form geschrieben wurde. Die Untersuchung deutet daraufhin, dass die Verwendung der anderen Antwortformen die Wahrscheinlichkeit der Richtigkeit der Antwort steigen lässt.

Die Überprüfung der Beziehung zwischen einer gut bewerteten Antwort und dem Rang des Antwortenden soll Aufschluss über einen möglichen Zusammenhang zwischen einem hohen Rang und einer richtigen Antwort geben (vgl. Abb. 4.15).

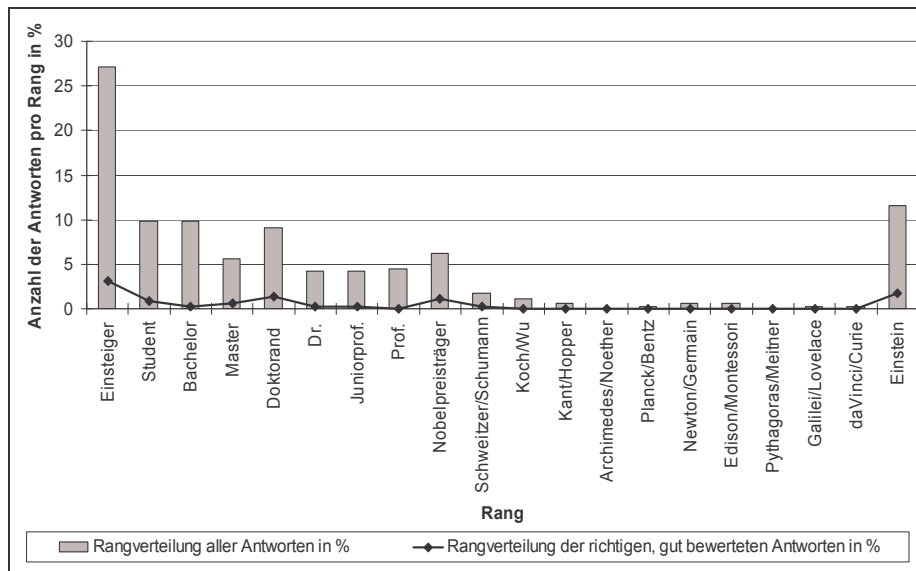


Abb. 4.15 Vergleich Rangverteilung gesamt (n=354) und Rangverteilung der richtigen, gut bewerteten Antworten (n=35) (Faktenfragen)

Die Grafik zeigt deutlich, dass kein Zusammenhang zwischen dem Rang eines Antwortenden und einer richtigen, gut bewerteten Antwort erkennbar ist. Stattdessen zeigt sich, dass bei der Rangverteilung der richtigen, gut bewerteten Antworten (Linie) eine ähnliche Verteilung wie bei der gesamten Rangverteilung aller untersuchten Antworten (Balken) zu erkennen ist. Darauf weisen unter anderem die Anteile der Ränge Einsteiger und Einstein hin. Die beiden Ränge stellen sowohl bei der Rangverteilung aller Antworten, als auch bei der Rangverteilung der richtigen, gut bewerteten Antworten die häufigsten Expertenstatus dar.

Von 16 als gut bewerteten Antworten, welche der Antwortende in seinem eigenen Themengebiet beantwortet hat, fielen zwölf Antworten (75 Prozent) richtig aus.

Es gilt noch zu klären, ob es einen Zusammenhang zwischen einer häufigen gut-Bewertung durch andere User und der Richtigkeit einer Antwort gibt. Dies wird bei der Untersuchung der Korrelation von der Häufigkeit einer gut-Bewertung und einer richtigen Antwort deutlich. Von insgesamt 44 gut bewerteten Antworten waren 35 richtig. Von diesen 35 wurden 26 Antworten (74,29 Prozent) jeweils einmal von anderen Usern als gut bewertet. Insgesamt sechs Antworten (17,14 Prozent) wurden jeweils zweimal und drei Antworten (8,57 Prozent) wurden jeweils dreimal als gut bewertet.

Es bleibt festzuhalten, dass hierbei nicht von einer Korrelation zwischen der Häufigkeit der gut-Bewertung und einer richtigen Antwort gesprochen werden kann.

3.3.1 Zusammenfassung der Auswertung des Faktenfragenkataloges

Bei der Auswertung der Faktenfragen können folgende allgemeine Ergebnisse festgehalten werden:

- Faktenfragen werden durchschnittlich mit drei Antworten bedacht, die erste Antwort erhält der Fragende in ca. drei Minuten.
- Um die Qualität der Antworten betrachten zu können, erfolgt eine Konzentration auf die Richtigkeit und die Art der Antwort. Bei den Faktenfragen sind die Antworten zu 80,2 Prozent korrekt, was eine solide Grundlage für einen Frage- Antwort Dienst darstellt.
- Der größte Anteil an Antworten (92 Prozent) wird mit mehr als zwei Worten beantwortet und stellt so in Folge der aufgestellten Definitionen einen Fließtext dar. Links werden nur in geringeren Mengen zur Beantwortung genutzt. Falls es zu einer Linkangabe kommt, wird in den meisten Fällen ein Fließtext genutzt, um den Link zu umschreiben oder anzukündigen.
- Betrachtet man die Bewertung der Faktenfragen, zeigt sich folgendes Bild: Ungefähr die Hälfte der untersuchten Faktenfragen wurde als hilfreich bewertet, gefolgt von den Top bewerteten und den unbewerteten Antworten. Die kleinste Einheit stellen die von andern Usern als gut bewertete Antworten dar.
- Bei der Betrachtung der unterschiedlich bewerteten Antworten (Top, hilfreich, gut von anderen Usern) wird deutlich, dass die höchste Anzahl an richtigen Antworten innerhalb des Bewertungsschemas „Top“ zu finden ist. Die Chance eine richtige Antwort zu finden sinkt mit den Kriterien „gute Bewertung durch andere User“ und „hilfreiche Antwort“

Deutet man diese Ergebnisse, stellt sich heraus, dass die Fragen bei Lycos iQ relativ schnell und in den meisten Fällen zufrieden stellend beantwortet werden. Auch die Nutzung von Wikipedia kann nicht übermäßig genannt werden. Lycos iQ stellt also keine „Brücke“ zu dem virtuellen Lexikon dar. Ebenfalls positiv kann die Bewertung der Antworten herausgestellt werden. Diese findet zum größten Teil statt und kann, durch die Tatsache, dass Top Antworten in den meisten Fällen korrekt sind, auch als funktionierend bezeichnet werden.

Es bleibt noch hinzuzufügen, dass ein Zusammenhang zwischen einer richtigen Antwort und einem bestimmten Rang nicht endgültig festgestellt werden kann. Allerdings weisen einige Grafiken daraufhin, dass Top Antworten häufiger von dem höchsten Rang (Einstein) gegeben werden. Da es sich bei Top Antworten in den meisten Fällen um richtige Antworten handelt, kann dies ein Hinweis auf die erwähnte Korrelation sein. Mit Sicherheit kann dies allerdings auf Grund der kleinen Fallzahlen nicht bestätigt werden.

3.4 Auswertung des Erfahrungsfragenkatalogs

Der Auswertung liegen 100 untersuchte Erfahrungsfragen zu Grunde. Auf diese 100 Erfahrungsfragen wurde insgesamt 393 Mal geantwortet. Aufgerundet wurde im Schnitt also auf jede Erfahrungsfrage vier Mal geantwortet.

Doch wie viele der gegebenen Antworten waren für die Fragenden brauchbar? Von 393 Antworten waren insgesamt 385 Antworten (97,96 Prozent) richtig. Die Menge der falsch beantworteten Antworten lag bei acht (2,04 Prozent). Das heißt also, dass sich acht Antworten nicht auf die jeweilige Frage bezogen haben.

Insgesamt ist allerdings zu beachten, dass 15 Antworten (3,82 Prozent) der insgesamt 393 Antworten Wort für Wort doppelt gegeben wurden.

Die durchschnittliche Zeit, die der User benötigt hat, der als erstes auf eine Frage geantwortet hat, liegt laut der Auswertung mit 132,25 Minuten sehr hoch. Dieses Ergebnis wurde von ein paar zeitlichen Ausreißern verzerrt. Durch den Median, welcher die oben genannten „Ausreißer“ ignoriert, ist zu errechnen, dass die durchschnittliche Antwortzeit ohne die zeitlichen Ausreißer bei vier Minuten liegt.

Es liegt im Interesse des Fragenden, eine möglichst umfangreiche Antwort zu bekommen. Entscheidend ist also, in welcher Form die Antworten gegeben werden. Die Kombinationen nur mit einem Link, mit Hilfe von Fließtext, mit Fließtext plus Link oder mit einem Wikipedia- Link zu antworten, wurden in unterschiedlicher Weise von den Antwortenden genutzt (vgl. Abb 4.1.6 Beantwortung allgemein Erfahrungsfragen (n=393)). Bei der folgenden Auflistung ist zu beachten, dass die gegebenen 393 Antworten 100 Prozent darstellen.

Von der ersten Form, dem ausschließlichen Beantworten durch einem Link, wurde acht Mal (2,04 Prozent) Gebrauch gemacht. Die zweite Form einer Fließtext-Antwort wurde 378 Mal (96,18 Prozent) verwandt. Die Anzahl der Antwortenden, die durch eine Kombination aus Fließtext und Link geantwortet haben liegt bei 30 (7,63 Prozent). Die vierte Form der Beantwortung durch die Unterstützung eines Wikipedia-Links wurde zwei Mal (0,51 Prozent) in Anspruch genommen. Die genaue Betrachtung dieser Werte gibt deutlich preis, dass die Antwort mittels eines Fließtextes die häufigste Form der Beantwortung war (vgl. Abb 4.16).

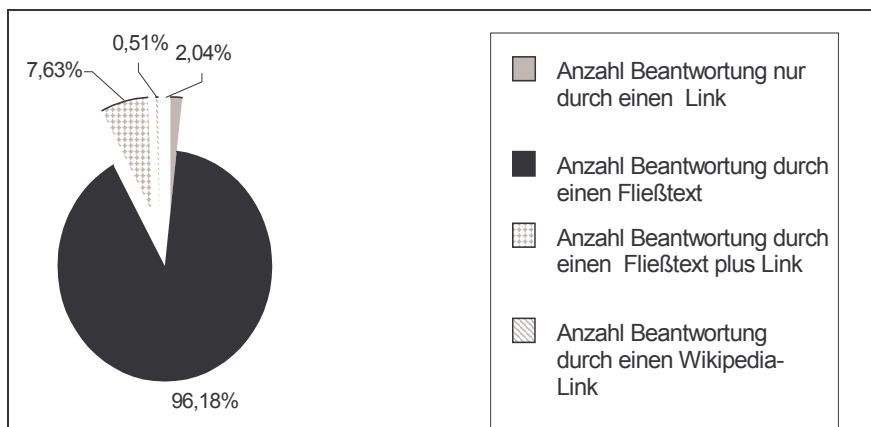


Abb. 4.16 Beantwortung allgemein Erfahrungsfragen (n=393)

Dies lässt zwei mögliche Begründungen zu. Zum einen ist es möglich, dass die Antwortenden entweder im Bereich ihres Themengebietes geantwortet haben und somit die Antwort ihrem Allgemeinwissen entnommen haben. Dies liegt nahe, da es sich hier um Erfahrungsfragen handelt, die, wie das Wort bereits beinhaltet, die Erfahrungen der Lycos iQ Nutzer abfragt. Zum anderen ist es möglich, dass die Antwort einer Quelle entnommen wurde, welche aber nicht genannt wird. Das geringe Vorkommen von Wikipedia-Links ist damit zu erklären, dass die Beantwortung von Erfahrungsfragen von den eigenen Erfahrungen, also subjektiv erlebten Ereignissen, abhängt. Die freie Enzyklopädie Wikipedia hingegen enthält laut eigenen Angaben

ausschließlich „bedeutsames Wissen aus belegten und zuverlässigen Quellen“ [1], was Erlebnisberichte ausschließt.

Die zweithäufigste Form der Beantwortung war die Kombination Fließtext und Link. Genau wie bei der Untersuchung der Faktenfragen lässt auch dieses Ergebnis darauf schließen, dass dies wie folgt zu begründen ist. Zum einen ist es möglich, dass die Antwortenden ihrer Antwort eine Quellenangabe hinzugefügt haben, um eine formal richtige Antwort zu geben, oder der Antwortende wollte jedem Leser mit dem Link die Entscheidung überlassen, ob er sich nachträglich durch die so angegebenen Informationen über das Thema informieren möchte. Beide Erklärungen zeigen auf, dass es sinnvoll ist, sich als Antwortender für diese Variante zu entscheiden. Es ist denkbar, dass User die diese Form wählen, wenn sie eine möglichst genaue Antwort geben möchten.

Die dritthäufigste Form der Beantwortung ist die der ausschließlichen Linkangabe. Hier besteht die Vermutung, genau wie bei den Faktenfragen, dass die Begründung für das geringe Vorkommen dieser Art der Antwort wieder mit dem „Geist“ der Community zu erklären ist. Die Antwortform der ausschließlichen Linkangabe verspricht die geringste Chance auf eine der drei positiven Bewertungen anderer User oder des Fragenden. Auch ist es nicht möglich, eine persönliche Note durch Wortwitz oder Emoticon, wie beispielsweise Smileys, einzubauen. Der virtuelle Charakter kann also auch hier gegenüber den anderen Community-Mitgliedern nicht positiv geformt werden. Der User, der sich für diese Form entscheidet, will möglicherweise lediglich schnell und unpersönlich sein Wissen an andere User weiter geben.

Die am wenigsten gewählte Form der Beantwortung ist die Antwort mittels eines Wikipedia-Links. In diesen Antworten sind allerdings auch Antworten enthalten, die auch in Kombination eines Fließtextes gegeben worden sind. Ebenso können es Antworten sein, welche ausschließlich durch einen Link beantwortet worden sind. Dieser Teil der Erfahrungsfragen-Untersuchung soll aufzeigen, wie oft die Antwortenden auf Wikipedia zurückgreifen und ob somit im Umkehrschluss eine große Anzahl der gegebenen Antworten ohne den Umweg über Lycos iQ auskommen würden. Das Ergebnis zeigt, dass sich diese Zahl aber lediglich knapp über einem halben Prozent bewegt. Diese Vermutung wurde also widerlegt.

3.4.1 Linkangabe in den Antworten

Ganz allgemein wurden 40 Antworten durch einen Link beantwortet. Diese Links sind von unterschiedlicher Qualität und wurden zur besseren Übersichtlichkeit in drei Kategorien eingeordnet, die bereits vorweg genannt wurden (vgl. Abb. 4.4). Die Einstufung in diese Kategorien sollte vor allem die Qualität der Quellen deutlich machen. Von den angegebenen 40 Links der Untersuchung waren zwei Links Wikipedia-Links. Bei der übrigen Menge von 38 Links ist allerdings ist zu beachten, dass von den Probanden nur 30 Links kategorisiert wurden. Es stammen sechs Links aus der Kategorie eins, zehn Links aus der Kategorie zwei und 14 Links aus der Kategorie drei.

Hier sieht man, dass die Angabe der inoffiziellen Links am häufigsten gewählt wurde. Dies lässt den Schluss zu, dass die angegebenen Links am häufigsten zu Quellen führen, welche keinen offiziellen Charakter haben. Bei diesen Quellen handelt es sich um Wikis, Blogs und private Homepages, welche möglicherweise

weiterführende Antworten auf die Erfahrungsfragen geben. Es liegt also nahe, dass die Antwortenden am häufigsten solche privaten Quellen gewählt haben, welche wahrscheinlich viel häufiger als Quellen von persönlichen Erfahrungen berichten. Es ist daher nicht als Manko zu betrachten, dass die User hier ihre Antworten durch eine inoffizielle Quelle gegeben haben.

Schlussendlich kann man auch bei den Erfahrungsfragen sagen, dass mit 46,67 Prozent knapp die Hälfte der Gesamtanzahl der angegebenen Links aus unverlässlichen Quellen stammt. Die restlichen Links verteilen sich auf 20 Prozent, der Links aus Top Quellen und 33,33 Prozent der Links aus offiziellen Quellen. Diese Links sind Quellen mit hoher bis zufrieden stellender Verlässlichkeit. Zusammenaddiert ergeben diese eine Prozentzahl von 53,33 Prozent, was sogar eine etwas größere Summe als die der nicht verlässlichen Quellen mit 46,67 Prozent darstellt.

3.4.2 Bewertung abgegebener Antworten

Bei den Erfahrungsfragen gibt es dieselben Bewertungsmöglichkeiten, um als Fragender dem Antwortenden und auch den anderen Community-Mitgliedern deutlich zu machen, für wie nützlich eine abgegebene Antwort erachtet wird. Diese Bewertungen können mehrfach innerhalb aller abgegebenen Antworten einer Frage vergeben werden. Der Fragende kann auch hier wählen zwischen Top-Antwort, hilfreiche Antwort oder nicht hilfreiche Antwort. Zusätzlich steht es dem Fragenden offen die Antwort nicht zu bewerten (Status nicht bewertete Antworten). Auch die anderen Community-Mitglieder haben die Chance, die abgegebenen Antworten einer Frage durch ein „gut“ zu bewerten.

Die 393 abgegebenen Antworten wurden insgesamt 457 Mal bewertet. Die bewertete Zahl ist deshalb höher, weil es möglich ist, eine Antwort durch den Fragenden und ebenso von anderen Community-Mitgliedern bewerten zu lassen. Von den untersuchten 393 Antworten wurden 88 (22,39 Prozent) als Top Antwort bewertet. Insgesamt 212 von 393 Antworten (53,94 Prozent) wurden als hilfreich bewertet. Die Anzahl der richtigen Antworten, die weder als Top Antwort noch als hilfreiche Antwort bewertet wurden und die auch nicht von anderen Community-Mitgliedern bewertet wurden, beträgt 118 (30,03 Prozent). Somit wurden 39 Antworten (9,92 Prozent) von anderen Community-Mitgliedern als gut bewertet.

Es wurde von der Vermutung ausgegangen, dass die von anderen Usern als gut bewerteten Antworten auch gleichzeitig richtig sind. Die Auswertung zeigte, dass dies bei den Erfahrungsfragen auch ohne Ausnahme der Fall ist. Alle der 39 als gut bewerteten Antworten waren richtig. Diese Ergebnisse aus der Stichprobe deuten darauf hin, dass die Bewertung durch andere User mit der Richtigkeit einer Antwort gleich zu setzen ist.

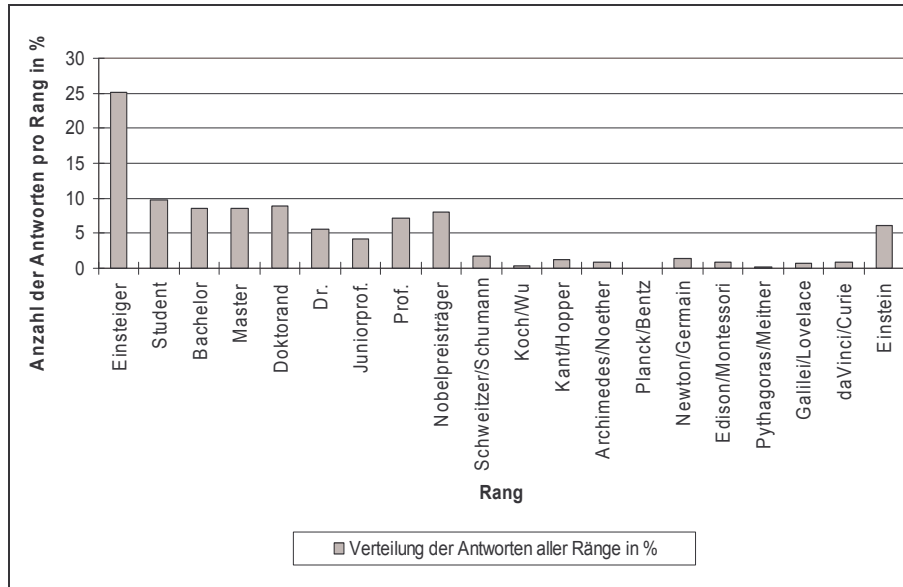


Abb. 4.17 Verteilung der Ränge der Erfahrungsfragen (n=457)

Bei Betrachtung der Rangverteilung innerhalb der Erfahrungsfragen (vgl. Abb. 4.17) zeigt sich ein ähnliches Bild wie bei der gesamten Rangverteilung der Faktenfragen. Es gibt viele Mitglieder mit dem Rang Einsteiger, die darauf folgenden Ränge sind weniger häufig vertreten. Die höheren Balken lassen sich durch die höhere Anzahl von Antworten erklären, welche bei Erfahrungsfragen gegeben wurden. Auffallend ist hier nur der Rang Einstein, welcher im Gegensatz zu den Faktenfragen nur einen geringen Anteil an der gesamten Rangverteilung einnimmt. Dies könnte darauf hindeuten, dass Mitglieder mit dem Rang Einstein eher Faktenfragen als Erfahrungsfragen beantworten.

3.4.3 Verteilung der Ränge der Antwortenden bei der am schnellsten abgegebenen Antwort

Die jeweils am schnellsten abgegebenen Antworten der 100 untersuchten Fragen wurden genau überprüft. Unter anderem wurde notiert, welchen Rang der Antwortende hat, welcher die jeweils schnellste Antwort gegeben hat. Es ist zu beachten, dass von den 100 untersuchten Usern vier keinen Rang innehatten bzw. dieser von den Probanden nicht angegeben wurde.

Bei der Auswertung war deutlich zu erkennen, dass von den 20 möglichen Rängen die User mit den Rängen Einsteiger, Student bzw. Doktorand und Master am häufigsten die schnellste Antwort abgegeben haben (vgl. Abb. 4.18). Dieses Ergebnis spiegelt deutlich die gesamte Rangverteilung wieder und spricht erneut dafür, dass die Ränge, welche in der größten Anzahl vorhanden sind, demnach auch die schnellsten Antworten geben. Dies liegt daran, dass die Wahrscheinlichkeit steigt, dass z.B. ein

User mit dem Rang Einsteiger auch der schnellste Antwortgeber ist, da einfach eine größere Masse an Einsteigern registriert ist.

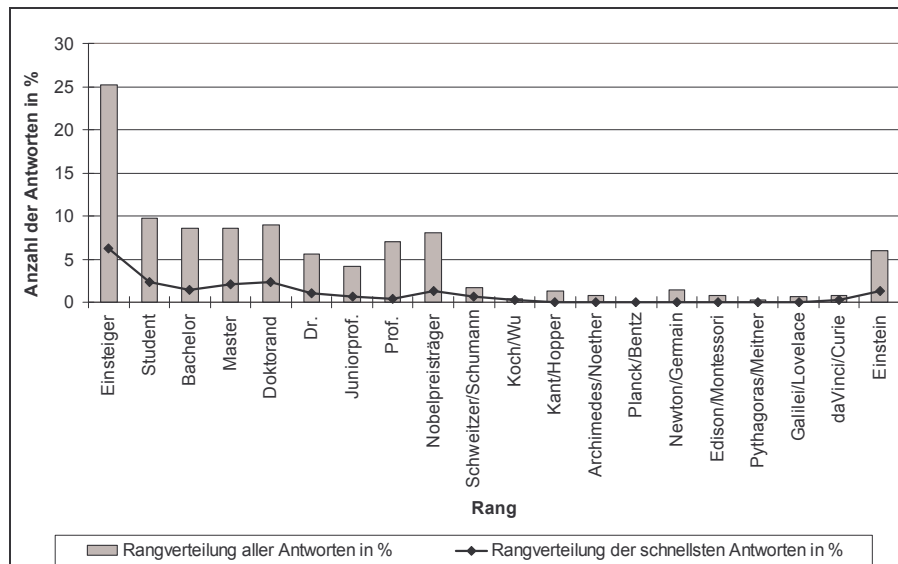


Abb. 4.18 Vergleich Rangverteilung gesamt (n=457) und Rangverteilung schnellste Antwort (n=96) (Erfahrungsfragen)

3.4.4 Untersuchung richtiger Antworten, die weder als Top, hilfreich oder gut bewertet wurden

Insgesamt 118 Antworten wurden weder als Top, hilfreich oder gut bewertet. Die Autoren dieser 118 Antworten konnten 20 mögliche Ränge innehaben. Auch hier zeigt sich ein ähnliches Bild wie bei der gesamten Verteilung der Ränge. Die meisten der richtigen, aber unbewerteten Antworten, wurden vom Rang Einsteiger gegeben, und die nachfolgenden Ränge gaben etwas weniger richtige, unbewertete Antworten. Es zeigt sich dennoch das gleiche Bild wie bei der Gesamtverteilung (vgl. Abb. 4.19).

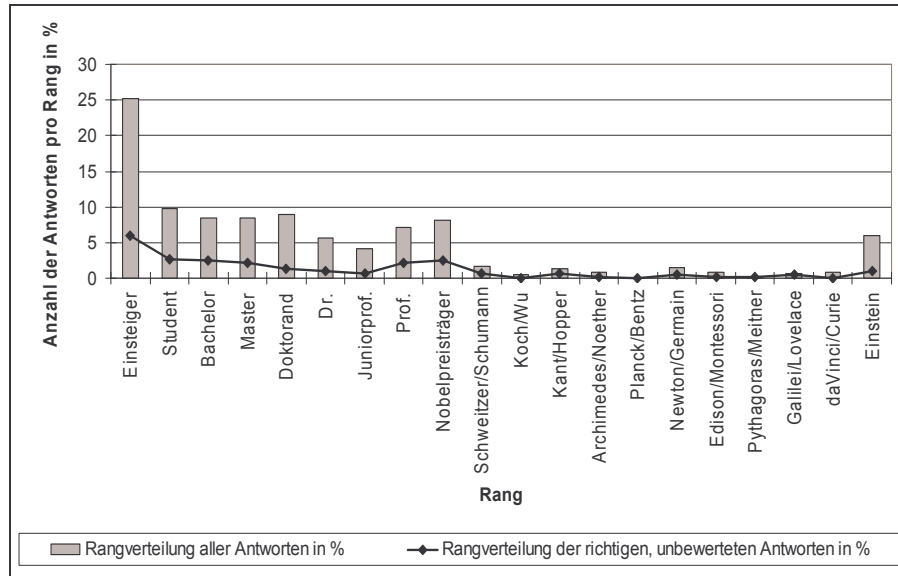


Abb. 4.19 Vergleich Rangverteilung gesamt (n=457) und Rangverteilung der richtigen unbewerteten Antworten (n=118) (Erfahrungsfragen)

118 Antworten wurden nicht bewertet, konnten aber durch Recherche als korrekt bezeichnet werden. Dabei hatten das jeweilige Themengebiet der Frage 41 (34,75 Prozent) User inne. 76 (64,41 Prozent) unbewertete Antworten wurden von Usern gegeben, die nicht in ihrem Themengebiet antworteten. Da das Profil eines Users zur Zeit der Auswertung deaktiviert war, konnte hier keine Untersuchung stattfinden.

Die Wahrscheinlichkeit ist groß, dass der Fragende entweder keine Bewertung abgegeben hat, weil es ihm zu aufwendig war, oder über diese Funktion gar nicht Bescheid wusste. Allerdings kann es auch sein, dass der Fragende keine der abgegebenen Antworten einer Bewertung für würdig fand.

Die genaue Untersuchung der Antworten, die weder als Top, hilfreich oder gut bewertet wurden, soll aufzeigen, in welcher Form geantwortet wurde. Von den 118 untersuchten Antworten wurden drei (2,56 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 113 Mal (96,58 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde neun Mal (7,69 Prozent) benutzt. Davon waren keine der angegebenen Links Wikipedia-Links. Hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Allerdings ergeben die Antworten, die ausschließlich durch einen Link gegeben wurden und die einen Fließtext plus Link enthielten, insgesamt eine Anzahl von zwölf Antworten, welche einen Link enthielten. Dies zeigt, dass die Vergabe von Links in den Antworten der Erfahrungsfragen eher die Ausnahme ist.

Die zwölf Links wurden noch einmal separat betrachtet und extra kategorisiert. Einer dieser Links konnte der ersten Kategorie zugeordnet werden, vier der zweiten

Kategorie und sieben der dritten. Diese Kategorisierung zeigt die Tendenz, dass auch hier die häufigste Quelle die „Inoffizielle Quelle“ ist.

3.4.5 Untersuchung der Top Antworten

Die Top Antworten haben eine Gesamtanzahl von 88. Von den 20 möglichen Rängen hatten die Antwortenden am häufigsten den Rang Einsteiger mit 21 Antworten (4,37 Prozent). Am zweithäufigsten hatten sie den Rang Doktorand mit zwölf Antworten (2,49 Prozent) und am dritthäufigsten den Rang Bachelor bzw. Master zu gleichen Teilen mit einer Anzahl von zehn Antworten (2,08 Prozent) inne. Auch hier wiederholt sich das Bild der Gesamtverteilung der Ränge. Alle 88 Fragen waren richtig.

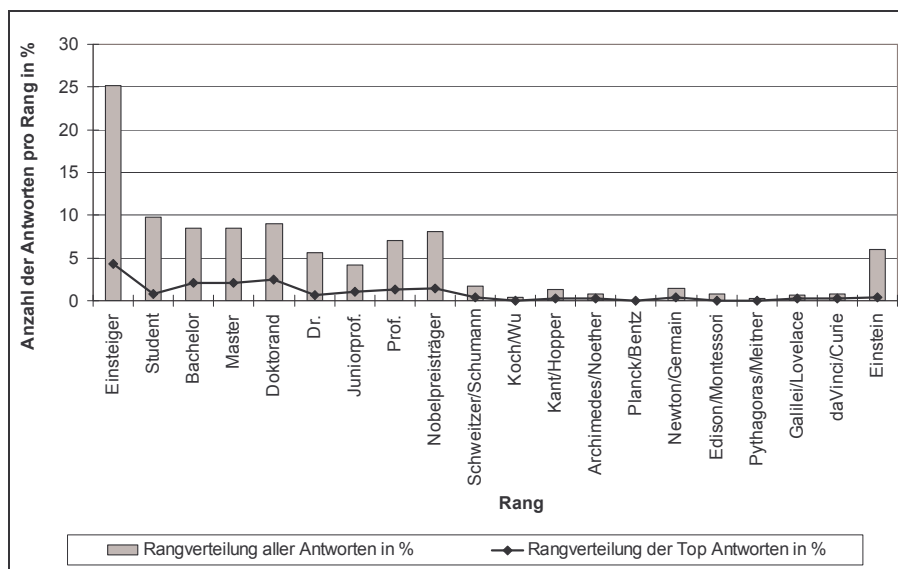


Abb. 4.20 Vergleich Rangverteilung gesamt (n=457) und Rangverteilung der Top bewerteten Antworten (n=88) (Erfahrungsfragen)

Es gab in dieser Kategorie insgesamt 27 (30,68 Prozent) Antworten, die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 60 Antworten (68,18 Prozent) wurden von Usern gegeben, die das jeweilige Themengebiet nicht als das Eigene angegeben haben. Bei einer Antwort (1,14 Prozent) konnte keine derartige Auswertung stattfinden, da der User mitsamt seinem Profil zurzeit der Auswertung bereits deaktiviert war; seine Antwort aber zur Ansicht noch aufrufbar ist.

Von den 88 untersuchten Top Antworten wurden drei (3,41 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 86 Mal (97,73 Prozent) benutzt und die Beantwortung durch einen Fließtext plus Angabe eines Links wurde acht Mal (9,09 Prozent) benutzt.

Davon war keiner der angegebenen Links ein Wikipedia-Link. Auch hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden und die, die einen Fließtext plus Link enthielten, ergaben insgesamt eine Anzahl von elf Antworten, welche einen Link enthielten. Dies zeigt erneut, dass bei den Erfahrungsfragen die Vergabe von Links in den Antworten eher die Ausnahme ist.

Die elf Links wurden noch einmal separat betrachtet und extra kategorisiert. Allerdings ist hier zu beachten, dass von den Probanden nur neun Links kategorisiert wurden. Es wurden zwei dieser Links aus der ersten Kategorie entnommen, ein Link wurde der zweiten Kategorie und sechs der dritten entnommen. Diese Kategorisierung zeigt die Tendenz, dass hier die häufigste Quelle der Links die „Inoffizielle Quelle“ ist.

Zusammenhänge zwischen der Richtigkeit einer Antwort und dem Rang des Antwortenden waren an dieser Stelle der Untersuchung nicht notwendig, da alle bewerteten Antworten auch gleichzeitig richtig waren.

3.4.6 Untersuchung der hilfreichen Antworten

Die hilfreichen Antworten haben eine Gesamtanzahl von 212. Die Autoren dieser 212 Antworten konnten 20 mögliche Ränge innehaben. Von diesen 20 möglichen Rängen hatten die Antwortenden am häufigsten den Rang Einsteiger mit 65 Antworten (13,51 Prozent). Am zweithäufigsten hatten sie den Rang Student mit 24 Antworten (4,99 Prozent) und am dritthäufigsten den Rang Doktorand mit einer Anzahl von 19 Antworten (3,95 Prozent). Es spiegelt sich also erneut die gesamte Rangverteilung wider (vgl. Abb. 4.21).

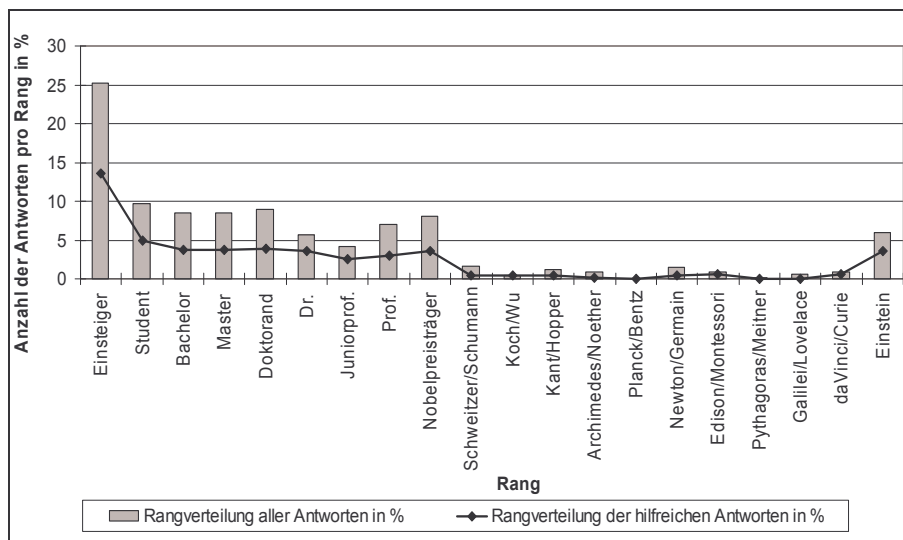


Abb. 4.21 Vergleich Rangverteilung gesamt (n=457) und Rangverteilung der als hilfreich bewerteten Antworten (n=212) (Erfahrungsfragen)

Bei der Untersuchung der hilfreichen Antworten hat sich ergeben, dass es durch die falsche Beantwortung der Probanden der Frage zwei im Erfahrungsfragenkatalog nun zwei verschiedene Ergebniszahlen vorliegen: zum einen das Ergebnis 212 Antworten und zum anderen 236 Antworten. Es wurde mit dem Ergebnis 236 Antworten weiter gearbeitet, da die Untersuchung der fehlerhaft beantworteten Frage nicht ohne größere Komplikationen noch einmal durchgeführt werden konnte. Insgesamt wurden die 236 Antworten von dem Fragenden als hilfreich bewertet. Davon waren alle 236 richtig und keine falsch.

Es gab in dieser Kategorie insgesamt 64 (27,12 Prozent) Antworten, die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 167 Antworten (70,76 Prozent) wurden von Usern gegeben, die das jeweilige Themengebiet nicht als das Eigene angegeben haben. Bei vier Antworten (1,69 Prozent) konnte keine derartige Auswertung stattfinden, da die User mitsamt ihrer Profile zur Zeit der Auswertung bereits gelöscht waren; ihre Antworten sind aber zur Ansicht noch aufrufbar. Bei einer Antwort (0,42 Prozent) wurde von den Probanden keine Bearbeitung vorgenommen.

Von den 212 untersuchten Antworten wurden drei (1,42 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 225 Mal (106,13 Prozent) benutzt und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde 18 Mal (8,49 Prozent) benutzt. Davon waren zwei (0,94 Prozent) der angegebenen Links Wikipedia-Links. Auch hier wird deutlich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden, die, die einen Fließtext plus Link enthielten und die Wikipedia-Link Antworten ergaben insgesamt eine Anzahl von 23 Antworten, welche einen Link enthielten.

Abzüglich der Wikipedia-Links sind es noch 21 Links, welche noch einmal separat betrachtet und kategorisiert werden. Es wurden vier Links aus der ersten Kategorie entnommen, acht Links wurden der zweiten Kategorie und neun Links der dritten entnommen. Diese Kategorisierung zeigt zwar, dass die häufigste Quelle der Links die „Inoffiziellen Quellen“ sind, zusammen genommen überwiegen allerdings die ersten beiden Kategorien, die die Quellen mit hoher Verlässlichkeit darstellen.

Zusammenhänge zwischen der Richtigkeit einer Antwort und dem Rang des Antwortenden waren an dieser Stelle der Untersuchung nicht notwendig, da alle bewerteten Antworten auch gleichzeitig richtig sind.

3.4.7 Untersuchung der Antworten, die von anderen Usern als gut bewertet wurden

Die als gut bewerteten Antworten haben eine Gesamtanzahl von 39. Die Antwortenden dieser 39 Antworten konnten 20 mögliche Ränge innehaben. Von diesen 20 Rängen hatten die Antwortenden am häufigsten den Rang Einsteiger, Student oder auch Doktorand zu gleichen Teilen mit sechs Antworten (1,25 Prozent) inne. Am zweithäufigsten nahmen sie den Rang Einstein mit fünf Antworten (1,04 Prozent) und am dritthäufigsten den Rang Master mit einer Anzahl von drei Antworten (0,62 Prozent) ein. Erneut zeigt sich in der Rangverteilung der als gut bewerteten Antworten die Gesamtverteilung der Ränge (vgl. Abb. 4.22). Alle 39 Antworten waren richtig.

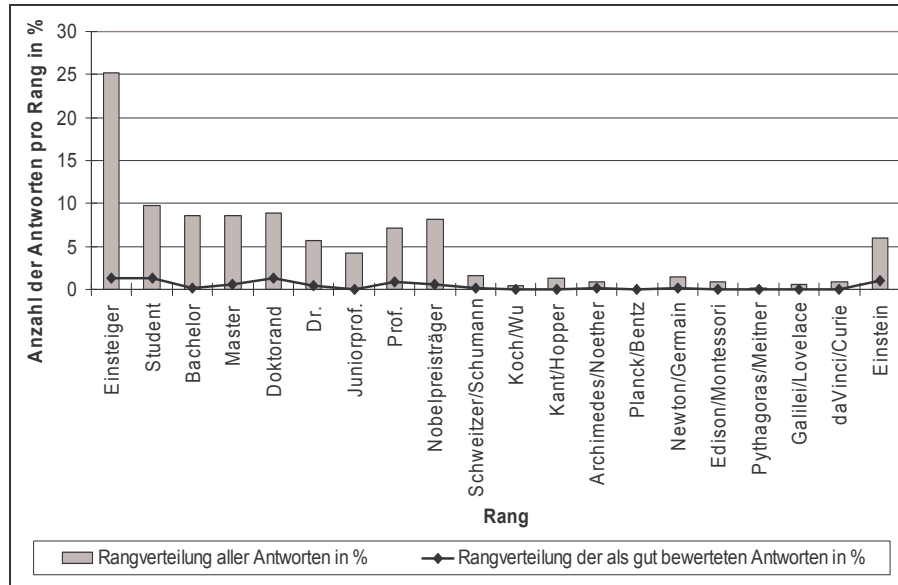


Abb. 4.22 Vergleich Rangverteilung gesamt (n=457) und Rangverteilung der gut bewerteten Antworten (n=39) (Erfahrungsfragen)

Es gab in dieser Kategorie insgesamt fünf Antworten (12,82 Prozent), die von Usern beantwortet wurden, die das jeweilige Themengebiet auch als eigenes Themengebiet angegeben haben. Eine Zahl von 34 Antworten (87,18 Prozent) wurden von Usern gegeben, die das jeweilige Themengebiet nicht als das Eigene angegeben haben.

Von den 39 untersuchten Antworten wurde eine (2,56 Prozent) ausschließlich durch die Angabe eines Links beantwortet. Die Beantwortung durch einen Fließtext wurde 38 Mal (97,44 Prozent) und die Beantwortung durch einen Fließtext plus die Angabe eines Links wurde ein Mal (2,56 Prozent) benutzt. Davon war keiner der angegebenen Links ein Wikipedia-Link. Es zeigt sich, dass die Antwort in Form eines Fließtextes am häufigsten gewählt wurde.

Die Antworten, welche ausschließlich durch einen Link gegeben wurden, und die, die einen Fließtext plus Link enthielten, ergaben insgesamt eine Anzahl von zwei Antworten, welche einen Link enthielten. Die Links wurden noch einmal separat betrachtet und extra kategorisiert. Allerdings ist hier zu beachten, dass von den Probanden nur ein Link kategorisiert wurde. Dieser Link entstammt der dritten Kategorie. Aufgrund der kleinen Fallzahlen können hier keine Tendenzen ausgemacht werden.

Eine Korrelation zwischen der Richtigkeit einer Antwort und dem Rang des Antwortenden waren an dieser Stelle der Untersuchung nicht notwendig, da alle bewerteten Antworten auch gleichzeitig richtig waren.

3.4.8 Zusammenfassung der Auswertung der Erfahrungsfragen

Bei einer erneuten Betrachtung der Ergebnisse der Untersuchung der Erfahrungsfragen fällt auf, dass die bewerteten Antworten zu 100 Prozent richtig sind. Dies erklärt sich durch die Eigenheiten einer Erfahrungsfrage. Bei diesem Fragetyp kann keine richtige Antwort ausgemacht werden. Vielmehr handelt es sich um Einschätzungen und Meinungen, die von dem Fragenden eingefordert werden.

Die allgemeinen Ergebnisse der Untersuchung zeigen, dass auf Erfahrungsfragen durchschnittlich vier Antworten gegeben werden. Die erste Antwort erreicht den Fragenden durchschnittlich in vier Minuten. Vergleicht man dies mit den Faktenfragen, zeigt sich dort genau eine Antwort weniger. Eine Interpretation dieses Aspektes weist darauf hin, dass Faktenfragen, welche genau einer richtigen Antwort bedürfen, daher mit weniger Antworten als Erfahrungsfragen auskommen. Antworten auf Erfahrungsfragen können durch eine viel größere Anzahl beantwortet werden, da hier grundsätzlich jeder User aufgefordert ist, seine Ansicht zu schildern.

Die Antworten sind fast alle im Stil eines Fließtextes gehalten. Links (auch Wikipedia- Links) werden nur zu verschwindend geringen Anteilen genutzt. Gründe für das Verhalten der Nutzer sind wieder in der „Natur“ der Erfahrungsfrage zu finden. Da die Ansichten der Nutzer abgefragt werden, scheint es auch wenig sinnvoll oder notwendig, Quellen zu zitieren. Hier ist das eigene Wissen gefragt.

Betrachtet man die Bewertung der Antworten so zeigt sich, ebenso wie bei den Faktenfragen, dass die größte Gruppe in den hilfreich bewerteten Antworten zu finden ist. Die folgende Reihenfolge weicht allerdings von der der Faktenfragen ab. Bei den Erfahrungsfragen folgen nun die unbewerteten Antworten und anschließend erst die Top bewerteten Antworten. Die Gruppe mit der geringsten Anzahl stellen hier die, von anderen Usern als gut bewerteten, Antworten dar.

Die große Anzahl an unbewerteten, aber trotzdem richtigen Antworten ist wieder einer Eigenschaft der Erfahrungsfragen zuzuschreiben. Bei einer Erfahrungsabfrage kann es zu Meinungsverschiedenheiten zwischen Fragendem und Antwortendem kommen, sobald unterschiedlich Ansichten bestehen. Der Konflikt äußert sich dann möglicherweise darin, dass der Fragende den Antwortenden gar nicht oder lediglich schlecht bewertet.

4. FAZIT

Die umfangreiche Auswertung führte zu einer Fülle von Ergebnissen und Interpretationen. Aufgrund dessen werden in diesem Teil der Auswertung ausschließlich die wichtigsten Ergebnisse und Interpretationen thesenartig zusammengefasst. Die Zusammenfassung erfolgt unter dem Blickwinkel der drei Ziele:

1. Feststellung der Qualität der Antworten
2. Herausarbeiten, ob der Expertenstatus mit qualitativ hochwertigen Antworten korreliert
3. Erarbeitung von Kriterien, um die Einbindung von qualitativen Antworten in die algorithmische Suche zu ermöglichen.

4.1 Die Qualität der Antworten

Um die Qualität der Antworten sinnvoll und übersichtlich darstellen zu können, erfolgt unter den folgenden Punkten eine Gegenüberstellung der Ergebnisse bei Fakten- und Erfahrungsfragen. Hier werden die essentiellen Merkmale der Antworten aufgezeigt und können einfach miteinander verglichen werden.

4.1.1 Allgemeine Untersuchungen

Allgemeine Kriterien	Faktenfragen	Erfahrungsfragen
Anzahl der untersuchten Fragen	100	100
Anzahl der gegebenen Antworten	293	393
Anzahl an Antworten pro Frage	3	4
Anteil richtiger Antworten	80,20%	97,96%
Anteil falscher Antworten	19,45%	2,04%
Minutenzahl bis erste Antwort	3 Min.	4 Min.
Anteil der doppelten Antworten	4,10%	3,82%
Anzahl an vergebenen Bonuspunkten	23	41
Anteil an Antworten Fließtext	92,15%	96,18%
Anteil an Antworten nur Link	3,07%	2,04%
Anteil an Antworten Fließtext + Link	22,87%	7,63%
Anteil an Antworten Wikipedia- Link	11,26%	0,51%
Bewertung der Antworten	Faktenfragen	Erfahrungsfragen
Top Antwort	32,42%	22,39%
hilfreiche Antwort	53,92%	53,94%
gute Antwort (nur durch andere User bewertet)	15,02%	9,92%
richtige Antworten (ohne Bewertung)	19,45%	30,03%

4.1.2 Untersuchung der Top Antworten

Allgemeine Kriterien	Faktenfragen	Erfahrungsfragen
Anzahl an Top Antworten	95	88
Anteil richtiger Antworten	92,63%	100%
Anteil falscher Antworten	7,37%	0%
Anteil an Antworten im eigenen Themengebiet	35,79%	30,68%
Anteil an Antworten Fließtext	88,42%	97,73%

Anteil an Antworten nur Link	6,32%	3,41%
Anteil an Antworten Fließtext + Link	35,79	9,09%
Anteil an Antworten Wikipedia-Link	12,63%	0,00%

4.1.3 Untersuchung der hilfreichen Antworten

Allgemeine Kriterien	Faktenfragen	Erfahrungsfragen
Anzahl an hilfreichen Antworten	158	212
Anteil richtiger Antworten	74,05%	100%
Anteil falscher Antworten	22,78%	0%
Anteil an Antworten im eigenen Themengebiet	28,01%	27,23%
Anteil an Antworten Fließtext	89,54%	100,00%
Anteil an Antworten nur Link	1,96%	1,42%
Anteil an Antworten Fließtext + Link	17,65%	8,49%
Anteil an Antworten Wikipedia-Link	8,50%	0,94%

4.1.4 Untersuchung der guten Antworten

Allgemeine Kriterien	Faktenfragen	Erfahrungsfragen
Anzahl an guten Antworten	44	39
Anteil richtiger Antworten	79,55%	100%
Anteil falscher Antworten	20,45%	0%
Anteil an Antworten im eigenen Themengebiet	36,36%	12,82%
Anteil an Antworten Fließtext	86,36%	97,44%
Anteil an Antworten nur Link	6,82%	2,56%
Anteil an Antworten Fließtext + Link	11,36%	2,56%
Anteil an Antworten Wikipedia-Link	2,27%	0%

4.1.5 Untersuchung der "richtigen" Antworten (die keine Bewertung erhielten)

Allgemeine Kriterien	Faktenfragen	Erfahrungsfragen
Anzahl an "richtigen" Antworten	57	118
Anteil an Antworten im eigenen Themengebiet	31,58%	34,75%
Anteil an Antworten Fließtext	77,19%	96,58%
Anteil an Antworten nur Link	3,51%	2,56%
Anteil an Antworten Fließtext + Link	24,56%	7,69%
Anteil an Antworten Wikipedia-Link	14,04%	0%

Zusammenfassend kann man sagen, dass die Qualität der Antworten mal subjektiv und mal objektiv zu betrachten ist. Eine ausschließlich objektive Bewertung war demnach nicht bei allen Teilen der Untersuchung möglich. Ein Beispiel für einen objektiv zu untersuchenden Bereich war die Richtigkeit der Faktenfragen-Antworten. Von den 293 abgegebenen Antworten waren ca. 80 Prozent richtig. Hier wurden die Antworten von den Probanden durch Recherche auf Richtigkeit geprüft.

Ein Beispiel für den entgegengesetzten Fall sind die Erfahrungsfrage-Antworten. Betrachtet man den Anteil an korrekten Antworten bei den Erfahrungsfragen kommt man auf eine höhere Prozentzahl von insgesamt 97,96 Prozent. Dies liegt an der Tatsache, dass alle Antworten, welche sich auf die Frage beziehen, als richtig eingestuft wurden

Daher ist es auch besonders schwierig an dieser Stelle über die Qualität der Antworten zu urteilen. Hier ist von den Probanden also lediglich eine subjektive Bewertung abgegeben worden. Da Lycos iQ allerdings nicht zwischen Faktenfragen und Erfahrungsfragen unterscheidet, kann man sagen, dass die Richtigkeit aller abgegebenen Antworten der durchgeführten Untersuchung zwischen aufgerundet 80 und 98 Prozent liegt.

Ein anderer Aspekt, welcher für die Qualität einer Antwort steht, ist die Form der Antwort. Die meisten Antworten wurden in einem Fließtext gegeben (mehr als zwei Wörter). Diese Tatsache könnte darauf hindeuten, dass es den Antwortenden wichtig war, dem Fragenden eine umfassende Antwort zu geben. Es wird vermutet, dass die Erklärung hierfür in der Motivation einer Community liegt.

Bei Lycos iQ werden Fragende wie Antwortende gleichermaßen durch das Bewertungssystem gefordert und motiviert aktiv an diesem Wissensaustausch teilzunehmen. Die Kommunikation durch das Bewertungssystem zwischen Fragendem und Antwortendem verhindert eine einseitige Community, in der der Reiz des Antwortgebens zweifelsohne geringer wäre. Der Antwortgeber kann so zu einem geschätzten Mitglied in der Community werden.

Diese These wird unterstützt von der Anzahl der User, die lediglich mit einem Link geantwortet haben, welche mit 3 Prozent sehr gering ist.

Außerdem wurde bei der Untersuchung der Faktenfragen deutlich, dass 23 Prozent der Antworten in Form eines Fließtextes plus Link gegeben wurden. Dies weist entweder darauf hin, dass den Antwortenden, die diese Form gewählt haben, eine Quellenangabe wichtig ist, oder sie die Intention hatten, auf diese Weise ihrer Antwort mehr Tiefe zu verleihen und es dem Fragenden zu ermöglichen, weitere Hintergrundinformationen einzusehen.

Beide Thesen untermauern die Untersuchungsergebnisse, welche zeigen, dass die meisten Antworten im Bereich der Faktenfragen von guter oder ausreichender Qualität sind. Es wird angenommen, dass die Antwortenden sich bemühen, ausführlich und korrekt zu antworten.

4.2 Zusammenhang zwischen Expertenstatus und Qualität der Antworten

Aufgrund der Untersuchung der einzelnen Ränge und deren Antworten, konnte der Frage nachgegangen werden, ob und inwieweit eventuell ein Zusammenhang zwischen der Höhe eines Ranges und der Qualität der Antworten besteht.

4.2.3 Faktenfragen

Die Auswertung der verschiedenen Ränge der Antwortenden im Bereich Faktenfragen machte deutlich, dass kein direkter Zusammenhang zwischen Rang und der Qualität der Antworten besteht. Es konnte lediglich eine leichte Korrelation zwischen richtigen Top Antworten und dem Rang Einstein festgestellt werden. Aufgrund der niedrigen Fallzahlen kann dies allerdings nicht mit Sicherheit bestätigt werden.

Eine qualitative Zunahme der Antworten vom Rang Einsteiger bis Einstein konnte also nicht sicher festgestellt werden. Vielmehr zeigte sich, dass untere Ränge insgesamt genauso qualitativ hochwertige Antworten lieferten, wie dies auch bei höheren Rängen beobachtet werden konnte. Insgesamt zeigt dies die hohe Motivation der freiwilligen Ratgeber, die, unabhängig von ihrem Rang, sich bemühen qualitativ gute Antworten zu geben.

Die hohe Quote von nahezu 80 Prozent richtigen Antworten im Bereich Faktenfragen untermauert diese These und zeigt, dass nahezu alle Ratgeber richtige Antworten produzieren wollen und dies tatsächlich auch tun.

4.2.4 Erfahrungsfragen

Da die Antworten der Erfahrungsfragen nach den Untersuchungen zu 97,97 Prozent „richtig“ waren, war eine gesonderte Prüfung eines Zusammenhangs zwischen Rang und Qualität der Antworten nicht durchführbar. Aufgrund der Subjektivität der Erfahrungsfragen konnte kein anderes messbares Kriterium, als der direkte Bezug der Antwort auf die Frage, herangezogen werden.

Da die Antworten sich nahezu immer auf die Fragen bezogen und nicht an diesen vorbei gingen, konnte es auch nicht vorkommen, dass zum Beispiel Einsteiger subjektive Erfahrungsfragen nach bereits genannter Definition „schlechter“ beantworteten als Antwortende mit dem Rang Einstein. Hier haben die Antwortgeber, unabhängig von den 20 unterschiedlichen Rängen, durchweg „richtige“ Antworten geliefert, so dass ein höherer Rang nicht mit qualitativ besseren Antworten korreliert hat.

4.3 Kriterien zur Einbindung der Antworten in die algorithmische Suche

Ein weiteres Ziel der Untersuchung war es, Kriterien zu erarbeiten, die eine Einbindung von qualitativ hochwertigen Antworten in die algorithmische Suche ermöglichen. Durch eine solche Einbindung würde der Suchende eventuell schneller eine treffende Antwort erhalten und Frage-Antwort-Dienste wie Lycos iQ würden außerdem verstärkt in das Bewusstsein der User rücken. Dies könnte einen erhöhten Zulauf bei Frage-Antwort-Diensten auslösen, welcher die Anzahl und Qualität der Antworten ebenfalls beeinflussen würde.

4.3.1 Faktenfragen

Eine Antwort sollte in die algorithmische Suche eingebunden werden, wenn sie einen Sachverhalt oder einen Fakt korrekt erläutert. Ein Kriterium zur Einbindung ist also die Richtigkeit einer Antwort.

Die Untersuchung ergab, dass die Top Bewertung einer Antwort die meisten richtigen Antworten kennzeichnet. Allerdings sinkt die Chance eine korrekte Antwort vorliegen zu haben mit den folgenden Bewertungsarten: Gute Bewertung durch andere User und hilfreiche Bewertung.

Das heißt also, dass wenn zu einer Suchanfrage eine Top bewertete Antwort vorhanden ist, sollte diese eingebunden werden. Die Wahrscheinlichkeit ist mit 92 Prozent enorm hoch, dass diese Antwort auch korrekt ist und den Suchenden zufrieden stellt.

Wenn jedoch keine Top Antwort vorhanden ist, dann sollte man eine Antwort einblenden, welche von anderen Usern als gut bewertet wurde. Hierbei spielt es keine Rolle, wie oft diese Antwort gut bewertet wurde.

Erst wenn auch hier keine Antwort vorhanden ist, sollte man eine als hilfreich bewertete Antwort in die Suche einblenden. In diesem Fall ist die Wahrscheinlichkeit allerdings am geringsten, dass es sich um eine korrekte Antwort handelt.

Man kann bei den Faktenfragen also die folgenden Kriterien zur Einbindung in die Suche mit einbeziehen, wobei darauf geachtet werden sollte, dass die Kriterien in ihrem Wert absteigen:

1. Top Bewertung
2. Gut-Bewertung durch andere User
3. Hilfreiche Bewertung.

4.3.2 Erfahrungsfragen

Da bei den Erfahrungsfragen nicht gefragt wurde, ob die Antwort richtig ist, sondern ob sie sich auf die Frage bezieht, sind hier fast alle Antworten korrekt gegeben worden. Bei der Untersuchung der einzelnen Bewertungsmöglichkeiten fällt keine falsche Antwort auf. Das heißt also, dass keinerlei Abstufung bei der Einbindung vorgegeben werden kann.

Der Umgang mit Erfahrungsfragen gestaltet sich deshalb als schwierig, weil eine Meinung oder Erfahrung nicht falsch sein kann. Aus diesem Grund kann jede oder keine Erfahrungsfrage in die algorithmische Suche mit eingebunden werden.

Die Empfehlung für die Einbindung in die algorithmische Suche ist daher, Fragen einzubinden, die die Kriterien der Faktenfragen erfüllen, denn eine Unterscheidung der verschiedenen Fragetypen kann nicht geleistet werden.

4.4 Allgemeine Anmerkungen

Eine der angegebenen Methoden konnte leider nicht mit Erfolg angewandt werden. Der Vergleich von Lycos iQ mit dem Frage-Antwort-Dienst Yahoo! Clever war nicht möglich, da dieses System komplett anders aufgebaut ist und der erarbeitete Kriterienkatalog hier keine Anwendung finden konnte. Zwar besteht bei Yahoo! Clever ebenfalls eine Art Bewertungssystem, allerdings können hier die User abstimmen, welche sie als die Beste Antwort ansehen oder aber, der Fragende entscheidet selbst über die beste Antwort. Es liegt allerdings keine Abstufung der Bewertung der Antworten vor. Außerdem gibt es keinen mit Lycos iQ vergleichbaren "Expertenstatus" oder Rang. Stattdessen können die User in sieben verschiedene Levels aufsteigen (siehe Kapitel 1: „Vergleich der Systeme/Marktanalyse“). Auch das

freie Tagging wird bei Yahoo! Clever nicht genutzt. Stattdessen sind die User angehalten, ihre Fragen in vorbestimmte Kategorien einzufügen.

Auf Grund dieser und weiterer Differenzen konnten jeweils nur ganz allgemeine Aspekte wie die Anzahl der erbrachten Antworten auf eine Frage oder die Richtigkeit der Antworten betrachtet werden.

Als Schwachpunkt von Lycos iQ fiel während der Bearbeitung die breite Fächerung der Themen des Frage-Antwort-Dienstes auf (über 220.000 Themen). Eine Reduzierung der Themen auf eine überschaubare Zahl würde die Übersicht für den Nutzer des Frage-Antwort-Dienstes verbessern (vgl den letzten Beitrag in diesem Band „Verbesserung des Taggings in Lycos iQ“).

Verwendete Literatur

1. Wikipedia (2008): Belege. Onlinedokument:
<http://de.wikipedia.org/wiki/Wikipedia:Belege> [Abruf am 25.01.2008]
2. Maaß, Christian/Gräfe, Gernot (2007): Alternative Searching Services: Seven Theses on the Importance of “Social Bookmarking”. Onlinedokument:
<http://www.christianmaass.com/wp-content/uploads/2008/03/maass-graefe-hess.pdf>
[Abruf am 07.07.2008]

Analyse der Fragensituation bei Lycos iQ

PHILIPP NIEHUES
philipp_niehues@gmx.de

DIANA SCHULZ
dischu77@gmx.de

VANESSA SCHÄFER
vschaefer@web.de

BIANCA OMELAN
bomelan@googlemail.com

Im nachfolgenden Bericht wird die Fragequalität auf der Plattform des Frage-Antwort-Dienstes Lycos iQ untersucht. Ziel ist es herauszufinden, welche Fragen qualitativ geeignet sind, um in algorithmische Suchmaschinen mit eingebunden werden zu können. Es stellte sich hierbei die Frage, ob sich aus der Qualität der Antworten Rückschlüsse auf die Qualität der Fragen ziehen lassen. Hierfür wurde ein Antworten-Kriterienkatalog entwickelt, der eine subjektive Bewertung der Antworten zuließ. Anschließend wurden die dazugehörigen Fragen im Hinblick auf inhaltliche und formale Kriterien beleuchtet. Dabei wurden außerdem eine Kategorisierung nach Frageart und eine Verteilung der Fragen nach Themengebieten vorgenommen. Korrelationen zwischen Antwort- und Fragequalität konnten festgestellt werden. Aus der genauen Analyse ergibt sich unter anderem, dass sowohl Faktenfragen qualitativ gute Antworten generieren als auch dass die Antwortqualität von Fragen aus Themengebieten mit allgemeinem, freizeitlichen Interesse am besten ist und sich daher für eine Einbindung in eine Suche eignen. Der Bericht schließt mit einigen Empfehlungen zur Erhöhung der Frage- und damit auch Antwortqualität.

1. EINLEITUNG

Der Erfolg von Lycos iQ hängt sicherlich zu großen Teilen auch davon ab, wie hilfreich die Antworten auf die gestellten Fragen von Gästen und Community-Mitgliedern sind. Gute Antworten sind zum einen für die Mitglieder der Community besonders hilfreich und damit ein Anreiz, sich weiter auf der Website zu engagieren, zum zweiten eignen sich sicherlich nur richtige und gute Antworten zur Einbindung in andere Suchformen. Wird also die Frage gestellt, wie Frage-Antwort-Dienste algorithmische Suchen ergänzen können, kommt man im nächsten Schritt zu der Frage, welche Antworten und dazugehörige Fragen gut genug sind, eine algorithmische Suche zu ergänzen, und wie diese guten Frage-Antwort-Paare aus der Masse der Daten zu extrahieren sind.

Der Ansatz dieses Beitrages ist dabei folgender: Wenn ein User in einer klassischen Suche eine Frage formuliert (in diesem Zusammenhang wird die Eingabe eines einzelnen Suchbegriffes ebenfalls als stark verkürzte Frage angesehen), so ist eine Beantwortung dieser Frage mit Hilfe der Antworten aus den Frage-Antwort-

Diensten erstmal nur dann möglich, wenn man die passenden Antworten zu der Suchanfrage auch finden kann. Hierzu wäre es sinnvoll, nach Übereinstimmungen zwischen der gesamten Suchanfrage des Users und den Fragen innerhalb von Lycos iQ zu suchen - und nicht nach einem bestimmten Stichwort innerhalb der gesamten Datenbasis - und diese passenden Fragen dem User dann als Ergänzung zu seiner Suchanfrage anzubieten.

Dieses Vorgehen setzt allerdings einen Zusammenhang zwischen der Qualität der Frage und der Qualität der zugehörigen Antworten voraus, da dem User wie bereits erwähnt ja nur qualitativ hochwertige Antworten weiterhelfen und man von der Ausgestaltung der Frage auf die Qualität der Antworten schließen können müsste.

Weiterhin müsste auch die Frage selbst qualitativ hochwertig formuliert sein, damit sie zum Beispiel vom Suchenden auch direkt als verständliche, nützliche Frage identifiziert werden kann.

Ziel dieses Beitrages ist es daher, Zusammenhänge zwischen der Qualität der Fragen und der Qualität der zugehörigen Antworten aufzuzeigen und Kriterien zu entwickeln, nach denen gute Fragen, also gut formulierte Fragen mit qualitativ hochwertigen Antworten, von den nicht geeigneten Fragen separiert werden können. Außerdem sollen anhand dieser Kriterien Vorschläge entwickelt werden, wie eine Anleitung des Users mit dem Ziel der Verbesserung der Qualität von Fragen und Antworten erreicht werden könnte.

2. THESEN

Als Grundlage für die oben genannten Überlegungen dienen folgende Thesen, die im Weiteren untersucht und verifiziert werden sollen:

- Zur Einbindung in andere Suchformen eignen sich nur Fragen mit guten und richtigen Antworten.
- Zur Einbindung in andere Suchformen eignen sich nur Fragen, die bestimmten qualitativen Merkmalen entsprechen.
- Es besteht eine Korrelation zwischen formalen und inhaltlichen Merkmalen einer Frage.
- Es besteht eine Korrelation zwischen formalen und inhaltlichen Merkmalen der Frage auf der einen Seite und der Qualität ihrer Antworten auf der anderen Seite.
- Geeignete Fragen lassen sich daher im Allgemeinen anhand bestimmter qualitativer Merkmale bestimmen.

3. METHODIK UND VORÜBERLEGUNGEN

3.1 Thema der Fragen

Die vorliegenden Ergebnisse stützen sich im Wesentlichen auf eine statistische Analyse der verfügbaren Datenbasis, das heißt auf eine Analyse einer bestimmten Anzahl von Fragen und Antworten mit Hilfe definierter Kriterien und einer anschließenden Auswertung der Ergebnisse. Hierzu wurde eine Stichprobe von 250 Fragen (247 vollständige Datensätze) mit dazugehörigen 393 Antworten ausgewählt

und von einem Probanden-Team ausgewertet. Dabei wurde darauf geachtet, dass bei der Auswahl der Stichprobe alle von der Arbeitsgruppe „Analyse der Antwort-Qualität“ definierten Themengebiete einigermaßen gleichmäßig abgedeckt wurden (vgl. Abb. 5.1). Eine genaue Beschreibung dieser Themengebiete findet sich im vorherigen Beitrag dieses Bandes „Untersuchung der Qualität der Antworten bei Lycos iQ und deren Einbindung in die algorithmische Suche“.

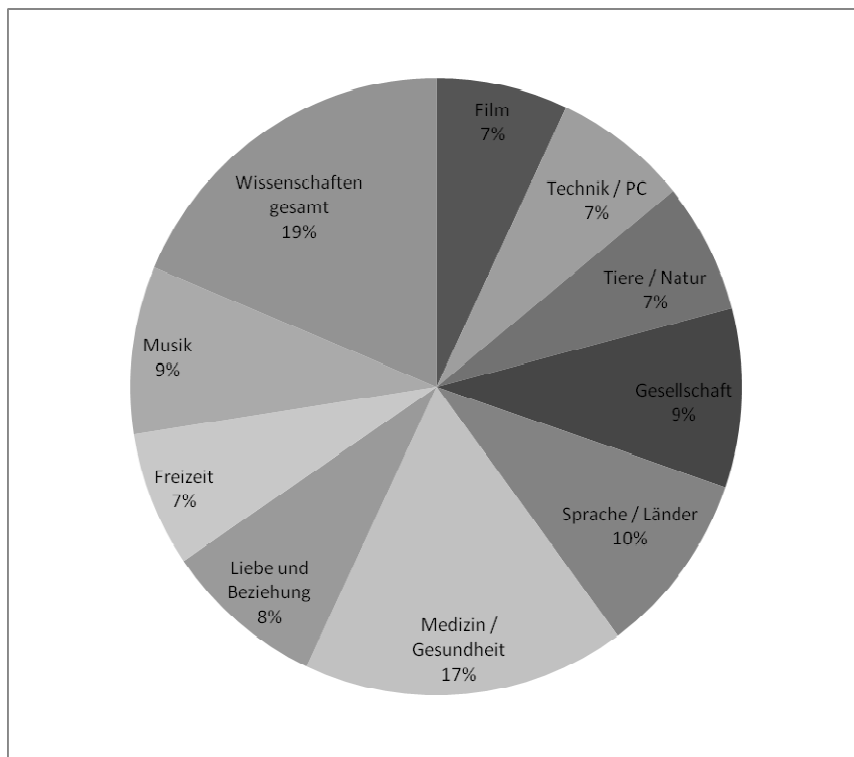


Abb. 5.1 Thematische Verteilung der Fragen

Die größere Anzahl der Fragen im Bereich Wissenschaften erklärt sich dadurch, dass hier verschiedene Wissenschaftliche Disziplinen (Politik, Geschichte, Literatur, etc.) zusammengefasst wurden.

3.2 Art der Fragen

Nach Auswahl der Fragen wurden diese nach ihrer Art kategorisiert. Bei der Einteilung lag die Annahme zu Grunde, eine bestimmte Frageart würde möglicherweise besonders qualitativ schlechte Antworten generieren, während eine andere Frageart dagegen mit auffällig qualitativ guten Antworten aufwarten könnte. Nach einer ersten Sichtung verschiedener zufällig gewählter Fragen zeigte sich, dass die Fragen in vier verschiedene Kategorien unterteilt werden können:

- **Faktenfragen**

Als Faktenfragen wurden alle Fragen definiert, auf die es auf jeden Fall eine eindeutig richtige Antwort geben muss, die in der Regel auch nachprüfbar ist.

Beispiele: Wo liegen die Osterinseln? Was ist eine Stammesfehde? Was ist der Unterschied zwischen einer GmbH und einer GmbH & Co KG?

- **Meinungsfragen**

Als Meinungsfragen galten alle Fragen, auf die es keine allgemein gültige Antwort geben kann. Die Fragen erlauben als Antwort eine subjektive Einschätzung und geben häufig die Möglichkeit zur Diskussion.

Beispiele: Gibt es Gott? Ist die Serie „Babylon 5“ empfehlenswert? Was ist besser...?

- **Lebensfragen**

Als Lebensfragen konnten Fragen bezeichnet werden, die als Antworten hilfreiche Tipps für den Alltag verlangen. Ob die Antworten auf solche Fragen wirklich sinnvoll sind und dem Fragesteller helfen, ist meist ohne entsprechenden Praxistest nicht direkt erkennbar.

Beispiele: Kann mir jemand ein günstigeres Mikro empfehlen? Wer kann mir da mal raten, wo ich Infos über ADHS herbekomme? Hat jemand Erfahrung mit...?

- **Spaßfragen**

Als Spaßfragen wurden Fragen bezeichnet, auf die der Fragesteller keine ernsthafte Antwort erwartet bzw. erwarten kann.

Beispiele: Warum gibt's hier bei Lycos iQ so viele Fragen? Sind alle Frauen dumm? Wie wird man immer glücklich?

Diese Einteilung der Fragen in vier verschiedene Arten basiert ebenfalls auf den Ergebnissen der Arbeitsgruppe „Analyse der Antwort-Qualität“, unterscheidet die Nicht-Faktenfragen aber noch differenzierter.

Um ein möglichst unverfälschtes Bild der quantitativen Verteilung der verschiedenen Fragearten zu bekommen, wurde eine zweite, größere Stichprobe ausgezählt. Für diese Stichprobe wurden 3259 Fragen ausgewertet und nach der Art der Frage kategorisiert. Dabei ergab sich die in Abbildung 5.2 dargestellte Verteilung:

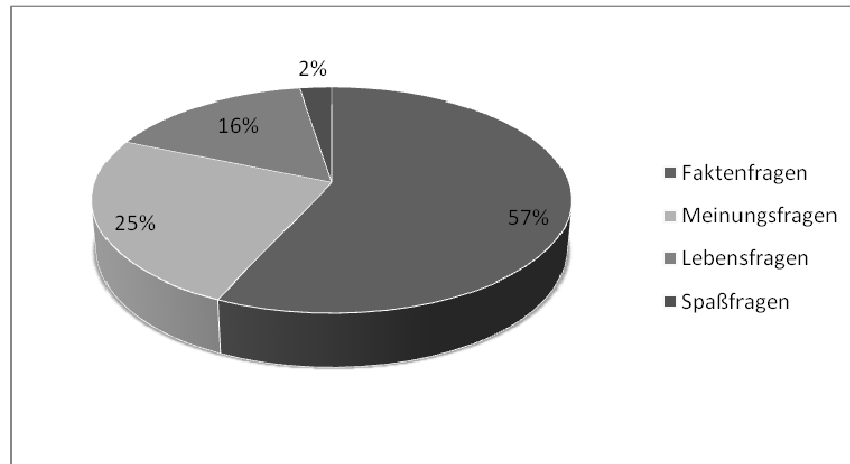


Abb. 5.2 Verteilung nach Art der Frage, große Stichprobe

Diese Verteilung stimmt allerdings bereits sehr gut mit der Verteilung innerhalb der kleineren Stichprobe überein (vgl. Abb. 5.3). Für die weiteren Analyseschritte werden daher die auf Grundlage der kleinen, 250 Fragen umfassenden Stichprobe erhobenen Daten verwendet.

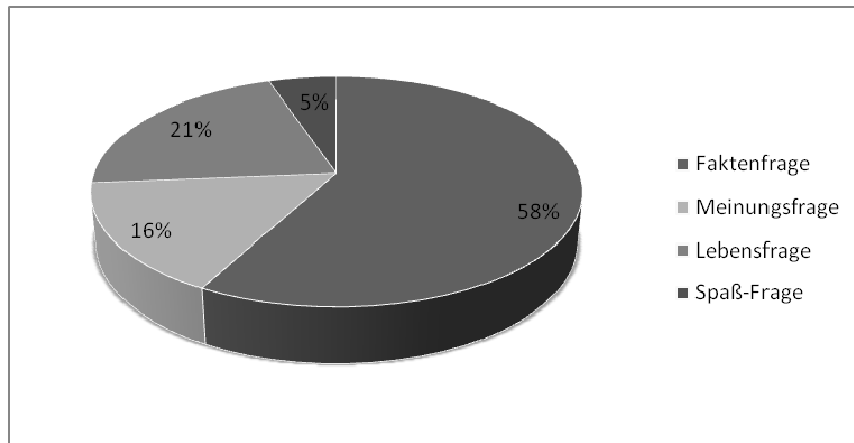


Abb. 5.3 Verteilung nach Art der Frage, kleine Stichprobe

Auffällig ist, dass mehr als die Hälfte aller Fragen Faktenfragen sind. Dies zeigt, dass es den Fragenden vor allem um eine konkrete Fragestellung geht. Auch der große Anteil der Lebensfragen, die meist nach einem Tipp zur Lösung eines bestimmten Problems suchen, bestätigt diesen Schluss. Obwohl gerade der Anteil der Meinungsfragen bei subjektiver Betrachtung manchmal höher erscheint, dient die Benutzung von Lycos iQ weniger dem Diskutieren kontroverser Themen oder dem Austausch von Meinungen zu einem bestimmten Thema.

Setzt man die Art der Frage in Abhängigkeit zu den verschiedenen Themengebieten, so ergibt sich die unten stehende Verteilung (vgl. Abb. 5.4). Das Ergebnis ist wenig überraschend, da der Anteil der Faktenfragen in Themengebieten, in denen eher eine Meinung als Fachwissen gefragt ist (z.B. „Liebe & Beziehung“, „Gesellschaft“, „Musik“), deutlich niedriger ist als in anderen Bereichen. Vor allem technische Themen wie „Technik & PC“ weisen einen sehr hohen Anteil an Faktenfragen auf. Dagegen fällt der Bereich „Wissenschaften“, dem man eigentlich einen recht hohen Anteil an Faktenfragen zubilligen würde, deutlich zurück, was damit begründet werden kann, dass hier verschiedenste wissenschaftliche Disziplinen zusammengefasst sind, die neben naturwissenschaftlichen auch gesellschaftliche oder politische Themen behandeln.

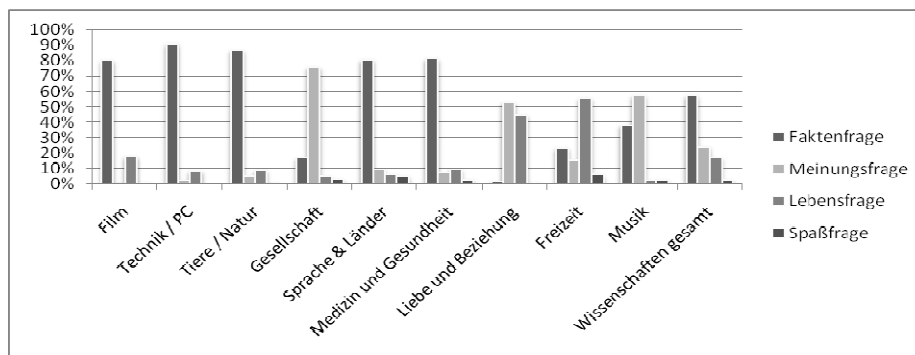


Abb. 5.4 Prozentualer Anteil der unterschiedlichen Fragearten in Abhängigkeit von den verschiedenen Themengebiete

4. ARBEITSCHRITTE

Die eigentliche Analyse gliederte sich in folgende Arbeitsschritte:

- **Erstellung eines Kriterienkatalogs zur Bewertung der Antworten**
Auf Grundlage der Ergebnisse der Arbeitsgruppe „Qualität der Antworten“ wurden die für die Qualität einer Antwort ausschlaggebenden Merkmale extrahiert und in Form eines Kriterienkataloges zusammengefasst.
- **Erarbeitung eines Punktesystems**
Um die Antworten bezüglich ihrer Qualität untereinander vergleichbar zu machen, wurde jedem Qualitätskriterium ein bestimmter Punktwert zugewiesen, der die Relevanz des Merkmales für die Qualität der Frage widerspiegelt.
- **Bewertung der Antworten**
Auf Grundlage des Punktesystems erfolgte eine Bewertung aller Antworten innerhalb der Stichprobe, so dass am Ende jeder Frage eine bestimmte Punktzahl als Indikator für die Qualität der Frage zugewiesen werden konnte.
- **Erarbeiten möglicher Qualitätskriterien für die Fragen**
Um die Qualität der Antworten in Relation zu den Fragen setzen zu können, wurde ein Kriterienkatalog entwickelt, der die Kategorisierung einer Frage nach bestimmten Merkmalen erlaubt.
- **Abbilden von Zusammenhängen**
Als zentraler Teil der Arbeit wurde im Anschluss versucht, Zusammenhänge zwischen einzelnen Frage-Kriterien und der Qualität der Antworten abzubilden, um so die für eine gute Frage ausschlaggebenden Merkmale zu extrahieren. Außerdem wurde versucht, Korrelationen der Merkmale untereinander herauszuarbeiten.
- **Formulierung von Empfehlungen**
Als letzter Teil wurden die Analyse-Ergebnisse auf die Fragestellung angewendet und in Form von Empfehlungen zur Einbindung von Fragen in anderen Suchformen und zur generellen Verbesserung der Fragequalität verdichtet.

Diese Arbeitsschritte werden im Folgenden eingehend erklärt.

5. KRITERIENKATALOG ZUR BEWERTUNG DER ANTWORTEN

Im Vorfeld musste zunächst geklärt werden, was eine qualitativ gute Antwort ist. Dafür wurden die Ergebnisse der Arbeitsgruppe zur Analyse der Antwort-Qualität herangezogen. Bei der Analyse der Antworten hat sich herauskristallisiert, welche Qualitätsmerkmale erfüllt sein müssen, um auf eine gute Antwort zu schließen. Im Folgenden werden die Merkmale einer guten Antwort ihrer Bedeutung nach (Rang) noch einmal kurz zusammengefasst, welche von der Arbeitsgruppe zur Analyse der Antwort-Qualität ermittelt wurden:

- Die Antwort muss richtig sein

- Die Antwort wurde mit Top vom Fragesteller bewertet, sie beinhaltet zusätzlich einen weiterführenden Link. Zudem hat der Antwortgeber in seinem eigenen Themengebiet geantwortet.
- Die Antwort wurde vom Fragesteller als hilfreich bewertet: Die Antwort zeichnet sich als Fließtext aus, beinhaltet zudem einen weiterführenden Link oder einen Wikipedia-Link.

Die Arbeitsgruppe zur Analyse der Antworten ermittelte ebenfalls die Ränge der Antwortgeber, die eine qualitativ gute Antwort gegeben hatten. Dieser Zusammenhang wird im Folgenden jedoch nicht berücksichtigt, da das Ergebnis verdeutlicht, dass ein „Einsteiger“ qualitativ bessere Antworten gibt als beispielsweise ein „Einstein“ oder „Bachelor“. Der Umkehrschluss muss demzufolge lauten, dass die Antwortgeber mit höherem Rang qualitativ schlechtere Antworten geben. Dies steht aber im Widerspruch zum Rang-Modell, denn der Antwortgeber hat mit aufsteigendem Rang im größeren Umfang qualitativ gute und richtige Antworten gegeben als die Antwortgeber der unteren Ränge. Aus diesem Grund wird der Rang des Antwortgebers nicht berücksichtigt.

5.1 Die Untersuchung der Antworten nach festgelegten Qualitätsmerkmalen

Das Analyse-Ergebnis der Arbeitsgruppe zur Analyse der Antworten lässt erkennen, welche Qualitätsmerkmale einer guten Antwort zugeordnet werden können. Die Arbeitsgruppe untersuchte jedoch die Qualität der Antworten ausschließlich im Hinblick auf die User-Bewertung. Es ist aber davon auszugehen, dass nicht alle User die Richtigkeit der Antworten zu ihrer gestellten Frage überprüfen. Deshalb musste eine Methode entwickelt werden, um die Qualität der Antworten subjektiv bewerten zu können.

Infolgedessen wurde eine Tabelle erstellt, welche die Qualitätskriterien der „Antworten-Arbeitsgruppe“ berücksichtigt, aber zudem eine Bewertung der einzelnen Antworten ermöglicht. Die Bewertung der Antworten wird mittels eines entworfenen Punktesystems vorgenommen. Jedem Kriterium wurden entsprechend der Bedeutung subjektiv beurteilt mehr oder weniger Punkte zugeteilt. Nachfolgend wird zunächst die Struktur der Tabelle vorgestellt (vgl. Abb. 5.5), auf Grund derer die Antworten untersucht wurden. Im Anschluss wird erläutert, wie die Antworten im Hinblick auf die festgelegten Qualitätskriterien bepunktet wurden.

Tabellenköpfe der einzelnen Spalten	Anmerkung
Fragen	Aufführung der Fragen aus der Stichprobe
Aus welchem Themengebiet stammt die Frage?	Die Unterteilung der Themengebiete in 10 grobe Kategorien erfolgte bereits durch die Antworten-Arbeitsgruppe. Nach deren Einteilung erfolgte die Zuweisung des Themengebietes.
Wie viele Antworten wurden gegeben?	Eintrag der Anzahl der gegebenen Antworten pro Frage
Wie hilfreich ist die Antwort?	Hier wurde eine subjektive Bewertung vorgenommen, inwiefern die Antworten tatsächlich hilfreich bei Lebens- und Erfahrungsfragen waren. Bei Faktenfragen wurde mittels Recherche geprüft, ob die Antworten tatsächlich richtig sind. Die Bepunktung dieses

	Qualitätskriteriums wird im nachfolgenden Punktesystem vorgestellt.
Wie viele Antworten sind richtig bzw. hilfreich?	Nur Eintrag der Antworten, die subjektiv gesehen tatsächlich hilfreich (bei Lebens- oder Erfahrungsfragen) oder richtig (bei Faktenfragen) sind.
Wie viele Antworten sind falsch, unverständlich oder unpassend?	Vornahme einer subjektiven Beurteilung: Bei Faktenfragen wurde die Anzahl der falschen Antworten eingetragen. Bei Lebens- und Erfahrungsfragen die Anzahl der Antworten, die nicht aussagekräftig oder unpassend sind.
Enthält die Antwort einen Fließtext?	Als Fließtext wird ein Text verstanden, der mehr als aus einem Wort besteht. Die einzutragende Anzahl der Antworten bezieht sich nur auf die Antworten, die zuvor subjektiv als hilfreich oder richtig bewertet wurden. Entsprechend wurden hier Punkte pro Antwort vergeben. Die Bepunktung wird nachfolgend vorgestellt.
Wie viele der hilfreichen oder richtigen Antworten bestehen ausschließlich aus einem Link?	Angabe der Anzahl von Antworten, welche die Frage lediglich durch einen weiterführenden Link beantwortet haben
Wie viele Antworten wurden durch die Angabe von Links unterstützt, dass heißt Fließtext und Link?	Eintrag der Antworten, die vom Fragesteller selbst formuliert wurden; ein weiterführender Link unterstützt die Antwort
Wie viele Antworten wurden insgesamt als Top bewertet?	Die Antworten wurden bezüglich der Beurteilung aus Fragestellersicht bewertet. Die dazu vorgenommene Bepunktung wird anschließend erläutert.
Wie viele Antworten wurden insgesamt als hilfreich bewertet?	Die Antworten wurden bezüglich der Beurteilung aus Fragestellersicht bewertet. Die Vergabe der Punkte wird anschließend erläutert.
Wie viele Antworten wurden weder als Top noch als hilfreich bewertet?	Der Fragesteller kann entweder eine Antwort nicht bewerten oder als nicht hilfreich bewerten. Hier sollten nur die Anzahl von Antworten eingetragen werden, die nicht vom Fragesteller bewertet wurden.
Wie viele Antworten wurden als nicht hilfreich bewertet?	Diese Antworten konnten in Lycos iQ durch einen Link eingeblendet werden. Die entsprechende Anzahl nicht hilfreicher Antworten sollte hier eingetragen werden.
Wie viele „gut“ Bewertungen bekamen die Antworten auf die Frage insgesamt?	Andere User (nicht der Fragesteller) können die gegebenen Antworten als gute Antworten bewerten. Wurde eine Antwort als gut bewertet, wurde eine Bepunktung vorgenommen, die anschließend erläutert wird.
Nach wie vielen Minuten kam die schnellste Antwort?	Eintrag der Anzahl der Minuten, nach der die erste richtige oder hilfreiche Antwort dem Fragesteller zugegangen ist. Dabei wurde anschließend eine Bepunktung vorgenommen, die im Anschluss vorgestellt wird.
Hat der Antwortgeber in seinem Themengebiet geantwortet?	Wenn der Antwortgeber, wie aus dem Lycos-Profil ersichtlich, in seinem angegebenen Themengebiet geantwortet hat, wurden entsprechend dem folgenden Punktesystem Punkte vergeben.
Zu wie viel Prozent hat der	Die bisher hilfreichen Antworten sind ebenfalls im Lycos-Profil des

Antwortgeber bisher hilfreiche Antworten gegeben?	Antwortgebers ersichtlich. Je nach prozentualem Anteil wurden nach dem Punktesystem entsprechend Punkte verteilt.
Summe Relevanz	Punkte aller Antworten, die subjektiv als hilfreich oder richtig bewertet wurden
Summe Fließtext	Punkte aller Antworten mit einem Fließtext
Summe Links	Punkte aller Antworten mit weiterführenden Links
Summe Bewertung	Punkte aller Antworten, die vom Fragesteller als Top oder hilfreich bewertet wurden oder wenn keine Bewertung vom Fragesteller erfolgte
Summe „gut“ Bewertung	Punkte aller Antworten, die von anderen Usern als gute Antwort beurteilt wurden
Summe Themengebiet	Punkte aller Antworten, bei denen der Antwortgeber nach seinem angegebenen Themengebiet geantwortet hat
Summe bisher hilfreiche Antworten des Antwortgebers	Punkte aller Antworten, die erreicht werden konnten, wenn der Antwortgeber bisher hilfreiche Antworten gegeben hat
Summe Geschwindigkeit	Punkte aller Antworten nach Beurteilung der Geschwindigkeit der ersten richtigen bzw. hilfreichen Antwort
Gesamtsumme	Gesamtpunktzahl aller erreichten Punkte in den einzelnen Qualitätskriterien
Wert pro Frage	Gesamtsumme dividiert durch die Anzahl der gegebenen Antworten zu einer Frage

Abb. 5.5 Untersuchungskriterien zur Bewertung der Antworten

5.2 Die Bewertung der Antworten nach den Qualitätskriterien anhand des Punktesystems

Die Beurteilung der Qualität der Antworten erfolgte subjektiv nach einem entwickelten Punktesystem. Die einzelnen Qualitätskriterien, die bezogen auf das Punktesystem bewertet wurden, werden nachfolgend noch einmal aufgeführt. Nicht jedem Qualitätskriterium wurde dieselbe maximale Punkteanzahl zugeordnet. Aus welchen Gründen hier differenziert wurde, wird im Anschluss erörtert.

5.2.1 Erstes Qualitätskriterium: Wie hilfreich ist die Antwort?

Punktesystem:

<i>sehr hilfreich</i>	= 50 Punkte pro Antwort
<i>bedingt hilfreich</i>	= 25 Punkte pro Antwort
<i>nicht hilfreich</i>	= 0 Punkte pro Antwort

Hier sollte überprüft werden, ob die Antwort für den Fragesteller hilfreich ist, wenn es sich um eine Erfahrungs- oder Lebensfrage handelt. Lag eine Faktenfrage vor, sollte die Richtigkeit mittels Recherche beurteilt werden.

Bei allen Fragearten wurde zudem beurteilt, ob eine sprachlich verständliche Formulierung der Antwort vorlag. Qualitativ war die Antwort auch bei allen Fragearten höher zu bewerten, wenn die Antwort eine Erklärung enthielt oder ausreichend begründet wurde.

Die Bepunktung erfolgte subjektiv, was auch notwendig war. Denn bei der Betrachtung der jeweiligen Antworten zu einer Frage wurde festgestellt, dass eine als Top, hilfreich oder gut beurteilte Frage durch den Fragesteller oder andere User nicht automatisch richtig oder hilfreich war.

Gerade bei Faktenfragen nimmt die richtige Beantwortung der Frage einen besonders großen Stellenwert ein, wenn davon auszugehen ist, dass diese Frage auf der Lycos-Suche zukünftig mit eingebunden werden soll. War die Antwort falsch, wurden daher keine Punkte vergeben.

Bei Lebens- oder Erfahrungsfragen wurde die Qualität der Antwort danach beurteilt, inwiefern diese hilfreich für den Fragesteller war. Eine Antwort war nur bedingt hilfreich beurteilt worden, wenn zwar ein Ratschlag gegeben wurde, jedoch nicht begründet wurde, warum dieser Ratschlag das Problem hilfreich lösen könnte.

Beispiel:

Wie bekomme ich einen Rotweinfleck aus meinem Teppich?

Antwort 1: Streu Salz drauf!

→ Ratschlag wurde zwar gegeben, aber keine Begründung oder Erklärung dazu.

Daher nur bedingt hilfreich = 25 Punkte.

Antwort 2: Wenn der Rotweinfleck frisch ist, streu sofort Salz darauf. Nicht damit sparsam sein, denn Salz zieht den Rotwein aus dem Teppich heraus.

→ Ratschlag wurde zudem auch begründet, daher sehr hilfreich = 50 Punkte.

Des Weiteren war die Antwort qualitativ höher, wenn der Antwortgeber die Beantwortung selbst formuliert hat und diese nicht nur durch einen angegebenen Link indirekt gegeben wurde.

Wie die Antwort-Arbeitsgruppe belegt hat, wird eine gute Antwort auch durch einen weiterführenden Link unterstützt. Aus diesem Grund wurden die angegebenen Links der Antwortgeber auf ihre Richtigkeit kontrolliert bzw. beleuchtet, inwiefern diese die Antwort hilfreich unterstützen. War das der Fall, konnte die Antwort zusätzliche Punkte erhalten (siehe Qualitätskriterium Nr. 2).

Bei dieser Bewertung konnte eine Antwort nach dem Punktesystem auch die höchste Punktzahl erreichen. Es ist davon auszugehen, dass der Fragesteller sich auf die Richtigkeit der Antworten verlässt, wenn es sich um Faktenfragen handelt. Eine besonders kritische Beurteilung dieser war daher unerlässlich.

5.2.2 Zweites Qualitätskriterium: Erhält die Antwort einen weiterführenden Link?

Punktesystem:

Ja = 10 Punkte pro Antwort

Nein = 0 Punkte

Wie bereits angemerkt, wurde die Qualität der weiterführenden Links unter Punkt 1 beurteilt. Wenn ein Link die Antwort hilfreich unterstützt hat, konnten zusätzlich 10 Punkte pro Link vergeben werden.

5.2.3 Drittes Qualitätskriterium: Erhält die Antwort einen Fließtext?

Punktesystem:

Ja = 10 Punkte

Nein = 0 Punkte

Eine Antwort besteht aus einem Fließtext, wenn sie aus mehr als einem Wort besteht. Die Qualität der Antwort erhöht sich, wenn diese auch für den Fragesteller begründet wird.

Beispiel:

Frage: Kennt jemand den St. Georgsorden?

Positiv: Ein Fließtext liegt vor, wenn der Antwortgeber begründet, wobei es sich bei diesem St. Georgsorden handelt.

Negativ: Wenn der Antwortgeber beispielsweise nur mit „Ja“ oder „Nein“ geantwortet hat.

5.2.4 Viertes Qualitätskriterium: Wie wurde die Antwort vom Fragesteller bewertet?

Punktesystem:

Top = 20 Punkte pro Antwort

Hilfreich = 10 Punkte pro Antwort

Keine Bewertung = 5 Punkte pro Antwort

Dieses Kriterium sollte der unter Punkt 1 vorgenommenen subjektiven Beurteilung entgegenstehen. Jedoch ist die maximale Punktzahl geringer als unter Punkt 1. Denn, wie bereits angemerkt wurde, war eine durch den Fragesteller als Top oder hilfreich bewertete Antwort nicht automatisch richtig. Wurde, wie das Punktesystem deutlich macht, die Antwort vom Fragesteller als Top bewertet, erhielt die Antwort die doppelte Punktzahl als eine hilfreiche Antwort. Eine Antwort konnte aber auch noch 5 Punkte dazu gewinnen, wenn sie überhaupt nicht vom Fragesteller bewertet wurde. Hintergrund dieser Punktevergabe ist, dass der Fragesteller eine

Antwort auch als nicht hilfreich beurteilen kann. Diese Antworten erhielten dann entsprechend keine Punkte.

5.2.5 Fünftes Qualitätskriterium: Wie oft wurde die Antwort von anderen Usern als „gut“ bewertet?

Punktesystem:

Für jede „Gut-Bewertung“ = 5 Punkte

Wenn eine Antwort von anderen Usern als „gut“ bewertet wurde, ist davon auszugehen, dass diese nicht nur für den Fragesteller qualitativ wertvoll ist. Die geringere Bepunktung liegt darin begründet, dass in erster Linie der Fragesteller mit der Antwort zufrieden gestellt sein sollte.

5.2.6 Sechstes Qualitätskriterium: Zu welchem Prozentsatz haben die Antwortenden bisher hilfreiche Antworten gegeben?

Punktesystem:

> 95 % = 10 Punkte

> 90 % = 5 Punkte

< 90 % = 0 Punkte

Bei diesem Kriterium wurde die Verlässlichkeit der bisherigen hilfreichen Antworten durch die Antwortgeber bewertet. Die bisher abgegebenen hilfreichen Antworten konnten im Lycos-Profil eingesehen werden. Hat der Antwortgeber bisher mehr als 95 Prozent aller Fragen hilfreich bzw. richtig beantwortet, wurden 10 Punkte vergeben. Lag die Quote zwischen 90 und 95 Prozent, so konnten 5 Punkte erreicht werden. Bei unter 90 Prozent der bisher hilfreichen bzw. richtigen Antworten wurden hingegen keine Punkte verteilt.

Zwar werden bei dieser Bepunktung die vergangenen gegebenen Antworten beurteilt, aber mit dieser zusätzlichen Punktevergabe differenziert sich heraus, in welchem Themengebiet besonderes viele richtige Antworten bisher gegeben wurden.

5.2.7 Siebtes Qualitätskriterium: Hat der Antwortende in seinem Themengebiet geantwortet?

Punktesystem:

Ja = 10 Punkte

Nein = 0 Punkte

Wie die Antwort-Arbeitsgruppe bei der Analyse der Antworten herausgefunden hat, sind Antworten von höherer Qualität, wenn der Antwortgeber aus seinem eigenen Themengebiet geantwortet hat, denn die Beantwortung im eigenen Themengebiet verstärkt die Glaubwürdigkeit der gegebenen Antwort. Daher wurden für jede

Antwort zusätzlich 10 Punkte vergeben, wenn der Antwortgeber entsprechend seinem Lycos-Profil in seinem Themengebiet geantwortet hat.

Eine höhere Punktevergabe wurde hier vernachlässigt, da die „Antworten-Gruppe“ herausgefunden hat, dass durchaus gute Antworten gegeben werden, wenn der Antwortgeber nicht im eigenen Themengebiet geantwortet hat.

5.2.8 *Achtes Qualitätskriterium: Dauer bis zur ersten richtigen Antwort*

Punktesystem:

<i>Schnell</i>	<i>< 60 Minuten = 10 Punkte</i>
<i>Mittel</i>	<i>< 400 Minuten = 5 Punkte</i>
<i>Langsam</i>	<i>> 400 Minuten = 0 Punkte</i>

Dieses Zeit-Kriterium wurde aufgenommen, weil davon auszugehen ist, dass eine Antwort für den Fragenden besonders hilfreich ist, wenn sie möglichst schnell gegeben wird. Die Einteilung der obigen Zeitbereiche wurde subjektiv vorgenommen und stützt sich auf die durchschnittliche Wartezeit bis zur ersten richtigen Antwort. In einigen Fällen kann eine erste richtige Antwort bereits nach 10 Minuten eintreffen, in anderen Fällen kam die erste richtige Antwort aber erst nach mehreren Stunden. Außerdem muss hierbei berücksichtigt werden, dass dem Antwortgeber auch die Möglichkeit gegeben sein sollte, sich ausführlich mit der Beantwortung einer Frage auseinanderzusetzen. Gerade im Hinblick auf die Erfahrungs- oder Lebensfragen ist die Antwort zumeist umfangreicher als bei einer Faktenfrage.

Qualitätskriterium	Punktesystem
Wie hilfreich ist die Antwort?	Sehr hilfreich = 50 Punkte Bedingt hilfreich = 25 Punkte Nicht hilfreich = 0 Punkte
Enthält die Antwort einen weiterführenden Link?	Ja = 10 Punkte Nein = 0 Punkte
Enthält die Antwort einen Fließtext?	Ja = 10 Punkte Nein = 0 Punkte
Wie wurde die Antwort vom Fragesteller bewertet?	Top = 20 Punkte Hilfreich = 10 Punkte Keine Bewertung = 5 Punkte
Wie oft wurde die Antwort von anderen Usern als „gut“ bewertet?	Für jede „gut“ Bewertung = 10 Punkte
Zu welchem Prozentsatz haben die Antwortenden bisher hilfreiche Antworten gegeben?	> 95 % = 10 Punkte > 90 % = 5 Punkte < 90 % = 0 Punkte
Hat der Antwortende in seinem Themengebiet geantwortet?	Ja = 5 Punkte Nein = 0 Punkte
Dauer bis zur ersten richtigen Antwort	Schnell < 60 Minuten = 10 Punkte Mittel < 400 Minuten = 5 Punkte Langsam > 400 Minuten = 0 Punkte

Abb. 5.6 Überblick des Punktesystems

6. METHODIK: FRAGEKATALOG

Nachdem 393 Antworten nach dem Punktesystem bewertet und qualitativ eingeordnet wurden, musste herausgefunden werden, ob sich aus der Art und Qualität einer Frage möglicherweise Rückschlüsse auf die Qualität der Antworten ergeben. Zu diesem Zweck war es erforderlich, die 247 Fragen, die zu den bereits bewerteten Antworten gehörten, näher zu betrachten und zu analysieren.

Für die systematische Untersuchung der Fragen war es notwendig, sich vorab Kriterien zu überlegen, mit deren Hilfe es theoretisch möglich sein könnte, die Qualität einer Frage zu beurteilen. Dabei lag bei den meisten Kriterien eine Hypothese zu Grunde, wie sich das gewählte Kriterium auf die Qualität der Antworten und Fragen auswirken könnte. Rückschlüsse, welches Kriterium sich tatsächlich für eine Unterscheidung der Fragen gut oder schlecht eignet, können aber erst nach Auswertung aller Daten getroffen werden.

Folgende Kriterien wurden vorab festgelegt:

6.1 Formale Kriterien

Es wurde davon ausgegangen, dass eine qualitativ gute Antwort von formalen Kriterien einer Frage abhängen könnte. Bei der Betrachtung formaler Merkmale sollte das inhaltliche Verständnis einer Frage unberücksichtigt bleiben. Es wurden folgende formale Kriterien, nach denen eine Frage bewertet werden kann, festgelegt:

- **Zeichnet sich die Frage durch eine gute Rechtschreibung aus?**

Jede Frage wurde auf Schwächen in der Rechtschreibung untersucht. Die Untersuchung ließ nur zwei Bewertungen zu: „Gut“ bedeutete, das Kriterium wurde beachtet, „Schlecht“ zeigte, das Kriterium wurde nicht erfüllt.

- **Wurde die Grammatik beachtet?**

Jede Fragen wurde auf Schwächen in der Grammatik untersucht. Die Untersuchung ließ nur zwei Bewertungen zu: „Gut“ bedeutete, das Kriterium wurde beachtet, „Schlecht“ zeigte, das Kriterium wurde nicht erfüllt.

Negativ-Beispiele: Wie entsteht eine Tsunami? Was ist von den Film "Lissi und der wilde Kaiser" zu halten? Gibt es schon Trailers? Welches ist die Hauptstadt von Mongolei?

- **Entspricht die Satzstellung einer Frage?**

Der Aufbau der Fragestellung wurde separat untersucht. Dabei sollte herausgefunden werden, ob eine Frage als solche richtig formuliert wurde. Es erschien wichtig, anhand der Satzstellung sofort erkennen zu können, dass es sich um eine konkrete Frage handelt. Das Kriterium ließ nur zwei Bewertungen zu: „Gut“ bedeutete, das Kriterium wurde beachtet, „Schlecht“ zeigte, das Kriterium wurde nicht erfüllt.

Beispiele für eine schlechte Fragestellung: Suche Infos zu Personen, Hauptstadt Boliviens, Igel zugelaufen, Definition für die Bereicherung.

6.2 Inhaltliche Kriterien

Es wurde vermutet, dass eine qualitativ gute Antwort von inhaltlichen Kriterien einer Frage abhängen könnte. Deshalb wurden die Fragen nach Merkmalen untersucht, die die inhaltliche Qualität einer Frage bewerten. Bei der Betrachtung dieser Merkmale sollte ausschließlich das inhaltliche Verständnis einer Frage berücksichtigt werden. Aus diesem Grund wurden folgende inhaltliche Kriterien festgelegt:

- **In welchem Sprachstil ist die Frage formuliert?**

Es wurde untersucht, ob die Frage in verständlichem Hochdeutsch formuliert oder ob Umgangssprache verwendet wurde. Es lag die Annahme zu Grunde, durch eine umgangssprachliche Formulierung der Frage würde es zu Missverständnissen kommen, welche falsche Antworten hervorrufen könnten.

Beispiel für eine Formulierung in Umgangssprache: Wie krieg ich's gebacken?

- **Werden Fremd- oder Fachwörter verwendet?**

Jede Frage wurde auf Verwendung von Fremdwörtern untersucht. Als Fremd- und Fachwörter sollten alle Wörter verstanden werden, die aus einer fremden Sprache oder aus einer Fachsprache übernommen wurden und deren Bedeutung im Allgemeinen nur schwer erfasst werden kann. Um Verständnis für die Frage zu erlangen, muss ein Antwortender die Fremd- und Fachwörter kennen oder die Bedeutung des spezifischen Begriffes nachschlagen. Es lag die Annahme zu Grunde, je weniger die Frage durch eine „schwere“ Sprache allgemein verständlich ist, desto schlechter könnten auch die Antworten ausfallen.

Beispiele: Wird Drive Defrag der TuneUp Utilities 2008 taugen? Was sind für euch die besten Psychedelic Hits?

- **Animiert die Fragestellung zu kurzen Antworten?**

Es wurde untersucht, ob die Fragestellung zu einer wenig ausführlichen Antwort animiert, obwohl davon auszugehen ist, dass der Fragesteller eigentlich mehr Informationen wünscht. Die Antworten könnten dann aus einem Wort bestehen wie „Ja“ oder „Nein“, „Gut“ oder „Schlecht“ und damit für den Fragesteller als unbefriedigend bzw. schlecht bewertet werden.

Beispiele: Kann mir jemand sagen, wo Senegal liegt? Glaubt ihr, es geht beim VfB wieder aufwärts in nächster Zeit? Wie findet ihr den Film ...?

- **Ist die Frage logisch verständlich?**

Es wurde untersucht, ob die Frage logisch verständlich formuliert wurde. Als logisch galt eine Frage, wenn aus ihr alle Informationen hervorgingen, die nötig waren, um eine eindeutige Antwort zu erhalten.

Beispiele für eine solche unlogische Frage: Kann mir jemand ein Buch empfehlen? Warum brennen immer meine Augen? Fabeltiere?

Um als logisch eingeordnet zu werden, musste die Frage außerdem eine inhaltliche Logik aufweisen. Es musste möglich sein, eine sinnvolle Antwort auf die Frage zu finden.

Beispiele für eine unlogische Frage: Gibt es Weihnachten 2020 Schnee? Was ist der mathematische Wert des Gottes?

- **Provoziert die Frage zu einer unpassenden Antwort?**

Es wurde untersucht, ob die Frage Antwortende dazu reizen könnte, eine ausfallende, verärgerte Antwort zu geben. Als provokativ galt eine Frage, wenn sie

Gelegenheit gab, Klischees zu bestätigen oder Tatsachen voraussetzte, die nicht vorauszusetzen waren.

Beispiele für provokative Fragen: Warum singen Schwarze besser als alle anderen? Warum sind Scientologen gefährlich? Wieso müssen sich Muslimas unter einem Kopftuch verstecken?

- **Wird die Ausführungsbox für weitere Erklärungen zur Frage genutzt?**
Es wurde untersucht, ob die Frage im anschließenden Zusatzfeld näher beschrieben wurde. Durch das Zusatzfeld wird dem Antwortenden die Möglichkeit gegeben, ein besseres Verständnis für den Sachverhalt zu erlangen, über den der Fragesteller etwas in Erfahrung bringen möchte. Es lag daher die Annahme zu Grunde, durch Ergänzungen im Zusatzfeld könnten bessere Antworten erzielt werden.
Beispiel?
- **Ist die Frage zeitlich begrenzt?**
Es wurde untersucht, ob die Frage einer zeitlichen Begrenzung unterliegt.
Beispiel: Wann findet das nächste Madonna-Konzert statt?

6.3 Schwierigkeitsgrad

Es erfolgte eine Einteilung nach dem Wissen, das ein Antwortender möglicherweise benötigt, um auch tatsächlich eine gute Antwort geben zu können. Der Wissensgrad wurde abhängig vom Rechercheaufkommen beurteilt. Es sollte folgende Annahme überprüft werden: Je speziellere Kenntnisse die Beantwortung einer Frage erfordert, desto schlechter könnten die Antworten ausfallen. Dabei wurde eine Einteilung in drei verschiedene Schwierigkeitsgrade vorgenommen: Allgemeinwissen, Fachwissen und Spezialwissen.

- **Kann ein Antwortender die Frage mit Allgemeinwissen beantworten?**
Es wurde untersucht, ob die gute Beantwortung der Frage nur allgemeines Wissen voraussetzt. Solche Fragen können von vielen Personen aus ihrem allgemeinen Kenntnisstand heraus oder aber durch eine schnelle und einfache Recherche beantwortet werden.
Beispiel für eine Frage, deren Beantwortung Allgemeinwissen benötigt: Wo steht das Empire State Building?
- **Kann nur ein Antwortender die Frage beantworten, der mindestens über Fachwissen verfügt?**
Es wurde untersucht, ob die gute Beantwortung der Frage Fachwissen voraussetzt. Solche Fragen können nicht von allen Personen aus ihrem allgemeinen Kenntnisstand heraus beantwortet werden. Ein Antwortender benötigt fachliche Kenntnisse über den Sachverhalt, muss sich mit dem Themengebiet schon im Vorfeld auseinandergesetzt haben oder für eine gute Antwort intensiv recherchieren.
Beispiel für eine Frage, deren Beantwortung Fachwissen benötigt: Hat Regina Halmig verdient gewonnen?
- **Kann nur ein Antwortender die Frage beantworten, der über Spezialwissen verfügt?**
Es wurde untersucht, ob die gute Beantwortung der Frage Spezialwissen voraussetzt. Solche Fragen können nur von wenigen Personen beantwortet werden,

die sich mit einer speziellen Thematik bereits ausgiebig beschäftigt haben. Die Fragen beinhalten oft bereits spezielle Fachwörter, über deren Bedeutung sich nur Antwortende mit speziellen Kenntnissen im Klaren sind.

*Beispiel für eine Frage, deren Beantwortung Spezialwissen benötigt:
Was ist besser, GoZilla oder FlashGet?*

6.4 Sonstige Kriterien

Nach der Feststellung formaler und inhaltlicher Kriterien der Fragen und der Berücksichtigung des Schwierigkeitsgrades zur Beantwortung einer Frage wurden sonstige weitere Kriterien festgelegt:

- **Wie viele Bonuspunkte wurden ausgesetzt?**

Es wurde untersucht, wie viele Bonuspunkte ein Fragesteller für die gute Beantwortung seiner Frage aussetzt. Dabei lag die Annahme zu Grunde, dass ein Fragesteller, der eine richtige und zufrieden stellende Antwort erwartet, eine höhere Anzahl Bonuspunkte als „Belohnung“ aussetzt als jemand, der es nur als Zeitvertreib und Spaß ansieht, Fragen bei Lycos iQ zu stellen. Des Weiteren gab es die Mutmaßung, ein Fragesteller, der Wert auf eine besonders gute Antwort legt und damit höhere Bonuspunkte aussetzt, würde sich gleichzeitig auch um eine formal und inhaltlich richtige Fragestellung bemühen.

- **Anzahl der Möglichkeiten, um eine Frage zu beantworten**

Es wurde untersucht, ob zu einer Frage mehrere richtige Antworten möglich sein könnten. Dabei lag die Annahme zu Grunde, dass bei Meinungs-, Lebens- und Erfahrungsfragen mehrere richtige Antworten vorstellbar sind – bei annähernd gleich bleibender Qualität. Bei Faktenfragen dagegen ist größtenteils nur eine richtige Antwort möglich, so dass dem Fragesteller doppelte und damit mehrfach dieselben richtigen Antworten zugehen. Diese Doppelungen könnten jedoch zu einer insgesamt schlechteren Beantwortung einer Frage führen, da ein Fragesteller, der bereits die eine richtige Antwort erhalten hat, die Antworten der "Nachahmer" wohl nicht mehr als nützlich empfinden würde. In einigen Fällen könnte es aber durchaus möglich sein, dass eine Faktenfrage mehrfach richtige Antworten generiert. Zum Beispiel, wenn es sich um eine allgemeine Personenfrage handelt, wie „Wer war Albert Schweitzer?“ könnten Antworten wie: "Theologe", "Orgelkünstler", "Musikforscher", "Philosoph" und "Arzt" in Frage kommen. In diesen Fällen könnte die unterschiedliche Beantwortung zu Erhöhung der Qualität der Antwort führen, da dem Fragesteller ein umfangreicheres Wissen zu seiner Frage vermittelt worden ist.

Kriterium	Untersuchung	Antwort
Einteilung nach Art der Frage:	Um was für eine Art Frage handelt es sich?	Faktenfrage
		Meinungsfrage
		Lebensfrage
		Spaßfrage
Formale Kriterien:	Zeichnet sich die Frage durch eine gute Rechtschreibung aus?	Ja/Nein
	Wurde die Grammatik beachtet?	Ja/Nein
	Entspricht die Satzstellung einer Frage?	Ja/Nein
Inhaltliche Kriterien:	In welchem Sprachstil ist die Frage formuliert?	Hochdeutsch/ Umgangssprache
	Werden Fremd- oder Fachwörter verwendet?	Ja/Nein
	Animiert die Fragestellung zu kurzen Antworten?	Ja/Nein
	Ist die Frage logisch verständlich?	Ja/Nein
	Provoziert die Frage zu einer unpassenden Antwort?	Ja/Nein
	Wird das Zusatzfeld für weitere Erklärungen zur Frage genutzt?	Ja/Nein
	Ist Frage zeitlich begrenzt?	Ja/Nein
	Wie viele User bewerteten die gestellte Frage als „Gute Frage“?	Anzahl der Gut-Bewertungen durch die User
Kriterien in Bezug auf den Fragesteller:	Wie viele Bonuspunkte wurden ausgesetzt?	Anzahl der Punkte
	Welchen Status hat der Fragesteller?	Status des Fragestellers
Einteilung nach Schwierigkeitsgrad:	Welches Wissen benötigt jemand, um die Frage zu beantworten?	Allgemeinwissen
		Fachwissen
		Spezialwissen

Abb. 5.7 Überblick aller Kriterien

7. METHODIK DER AUSWERTUNG

Die für die Analyse notwendigen Daten wurden manuell von einem Probanden-Team erhoben, indem Fragen und Antworten mit Hilfe der definierten Kriterien subjektiv beurteilt wurden. Anschließend erfolgte eine statistische Auswertung der Daten, um Zusammenhänge aufzuzeigen.

Da Zusammenhänge der einzelnen Antwort-Merkmale untereinander bereits im Beitrag der Arbeitsgruppe zur Analyse der Antwort-Qualität dargestellt wurden (vgl. den vorherigen Beitrag in diesem Bad), sollen hier nur die Kriterien erläutert werden, die in direktem Zusammenhang mit der jeweils zugehörigen Frage stehen. Als Grundlage dient also lediglich die nach oben erklärtem Muster errechnete Punktzahl pro Frage, die die Antwortqualität widerspiegelt.

8. ERGEBNISSE

8.1 Verteilung der Antwortqualität

Nach Auswertung aller 393 Antworten ergab sich eine durchschnittliche Punktzahl pro Frage von 70,09 Punkten. Eine Antwort durchschnittlicher Qualität würde also eine Punktzahl von 70,09 Punkten erreichen. Dabei waren maximal 120 Punkte zuzüglich je 5 Punkten pro vergebener Gut-Bewertung zu erreichen. Die in Abb. 5.9 dargestellte Verteilung ergibt sich aus der Ordnung der Fragen nach folgender Skala:

Sehr gut:	> 100
Gut:	100 < > 80
Mittel	80 < > 60
Schlecht	60 < > 40
Sehr schlecht	< 40

Abb. 5.8 zeigt exemplarisch die Punkteverteilung für eine Frage mit sehr guter Antwortqualität und eine sehr schlecht bewertete Frage.

Qualitätskriterien	Was bedeutet der Begriff „Halb-schmarotzer“ in der Pflanzenwelt?	Auf welche Beweise begründet sich die Behauptung, Jesus sei in einer Werkstatt geboren?
Wie hilfreich ist die Antwort?	100	25
Enthält die Antwort einen weiterführenden Link?	0	10
Enthält die Antwort einen Fließtext?	20	10
Wie wurde die Antwort vom Fragesteller bewertet?	40	0
Wie oft wurde die Antwort von anderen Usern als „gut“ bewertet?	10	15
Zu welchem Prozentsatz haben die Antwortenden bisher hilfreiche Antworten gegeben?	20	0
Hat der Antwortende in seinem Themengebiet geantwortet?	20	10
Dauer bis zur ersten richtigen Antwort	10	10
Summe	220	80
Anzahl der Antworten	2	5
Wert der Frage	110	16
Bewertung	Sehr gut	Sehr schlecht

Abb. 5.8 Exemplarische Bewertung einer Frage mit sehr guter und einer Frage mit sehr schlechter Antwortqualität

Abb.5.9 zeigt, dass ca. ein Drittel aller Fragen keine brauchbaren Antworten generieren. Angesichts der fehlenden qualitativen Kontrolle der Antworten in der Community ist dieser Wert positiv zu bewerten, vor allem da der Anteil der

mindestens guten Antworten deutlich höher ist als der Anteil der höchstens schlechten Antworten. Allerdings wurden auch nur 14 Prozent der Antworten von den Probanden als sehr gut und 23 Prozent als gut bewertet. Setzt man nun voraus, dass sich, wie eingangs erwähnt, ausschließlich gute und sehr gute Antworten zur Ergänzung klassischer Suchformen eignen, reduziert sich die Menge der potenziell geeigneten Antworten auf ca. ein Drittel der Gesamtmenge.

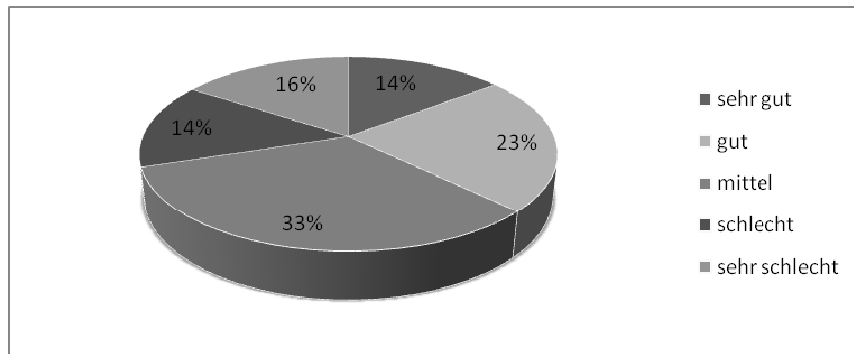


Abb. 5.9 Verteilung der Fragen nach ihrer Qualität

Betrachtet man die thematische Verteilung der Antwortqualität, so ergibt sich die in Abb. 5.10 dargestellte Verteilung. Auffällig sind hier vor allem die sehr gute Qualität der Antworten im Bereich „Filme“ und das gleichzeitig schlechte Abschneiden der Wissenschaftsbereiche. Hier liegt der Schluss nahe, dass Bereiche mit hohem allgemeinen Interesse (auch „Freizeit“, „Sprachen/Länder“) und niedrigem benötigten Fachwissen eine besonders gute Antwortqualität generieren. Symptomatisch für die spezialisierten Bereiche dagegen ist der Bereich „Computer & Technisches“, dessen Fragen, die in einem Technik-affinen Umfeld gestellt werden, nur eine durchschnittliche Antwortqualität von ca. 67 Punkten erreichten.

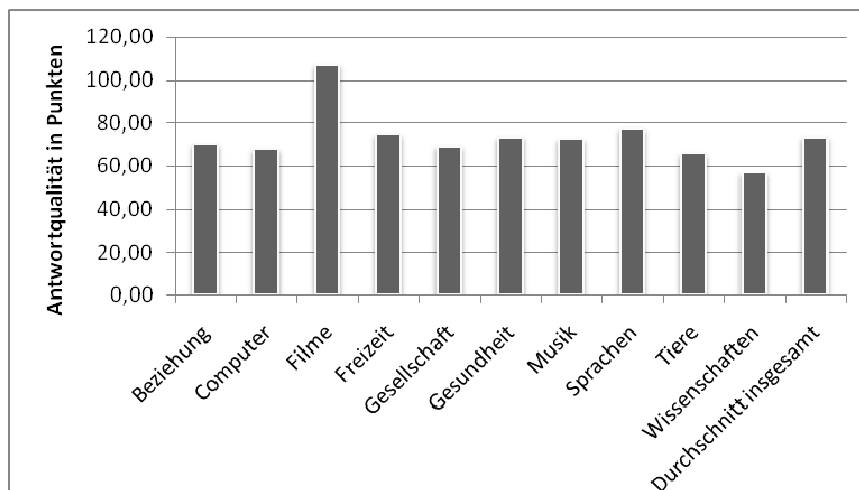


Abb. 5.10 Durchschnittliche Antwortqualität pro Themengebiet, n=247

Verteilt man die Antwortqualität nun auch nach der Art der Frage, erhält man eine wenig überraschende Verteilung, die den Faktenfragen die beste Antwortqualität zubilligt. Dies dürfte vor allem in der Art der Antworten begründet liegen, da Antworten auf Faktenfragen meist präziser und weniger strittig sind als Antworten, die eine einzelne Meinung widerspiegeln (vgl. Abb. 5.11).

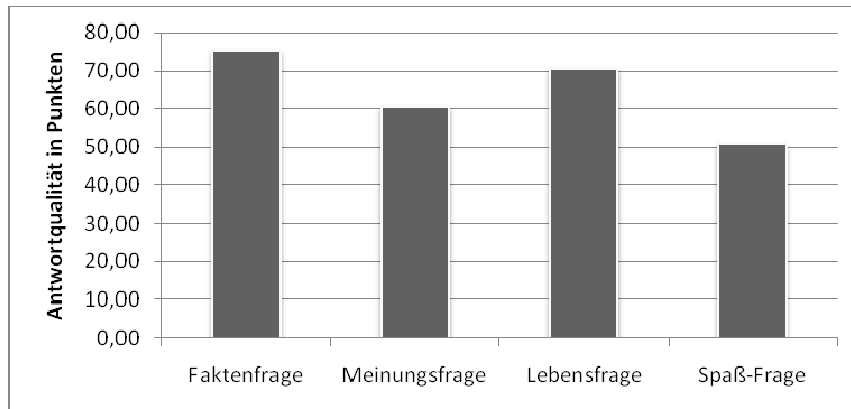


Abb. 5.11 Durchschnittliche Antwortqualität pro Art der Frage, n=247

Als erstes Ergebnis ist an dieser Stelle also festzuhalten, dass Fragen dann besonders gute Antworten generieren, wenn sie ernst gemeint sind und als Antwort keine Meinung verlangen. Themengebiete von hohem allgemeinen Interesse und eher geringem benötigten Fachwissen schneiden besser ab als schwierigere, spezielle Themen. Faktenfragen aus dem Bereich „Film“ erreichen so auch tatsächlich im Schnitt eine Antwort-Punktzahl von 86,08 Punkten, während Spaßfragen aus dem Bereich „Wissenschaften“ nur auf eine Punktzahl von 35,13 kommen.

Die Art der Frage und das Thema, in dem die Frage gestellt wurde, lassen also bereits erste Schlüsse über die zu erwartende Qualität der Antworten zu. Vor allem der Zusammenhang zwischen Antwortqualität und Thema der Frage ist dabei auf Grund der einfachen Möglichkeit einer späteren, automatisierten Selektion der guten Fragen interessant.

8.2 Schwierigkeitsgrad

Wenn man sich nun dem Zusammenhang zwischen der Qualität der Antworten und der analysierten Merkmale zuwendet, lässt sich das oben beschriebene Ergebnis auch mit der folgenden Grafik belegen (vgl. Abb. 5.12). Die Prozentzahlen spiegeln dabei jeweils den Anteil der Fragen mit positiver Bewertung wider. Wird der Anteil der Fragen, der mit Allgemeinwissen zu beantworten ist, in einem Themengebiet mit 90 Prozent angegeben, so bedeutet dies, dass die Frage: „Kann ein Antwortender die Frage mit Allgemeinwissen beantworten?“ in 90 Prozent der Fälle mit „Ja“ beantwortet wurde.

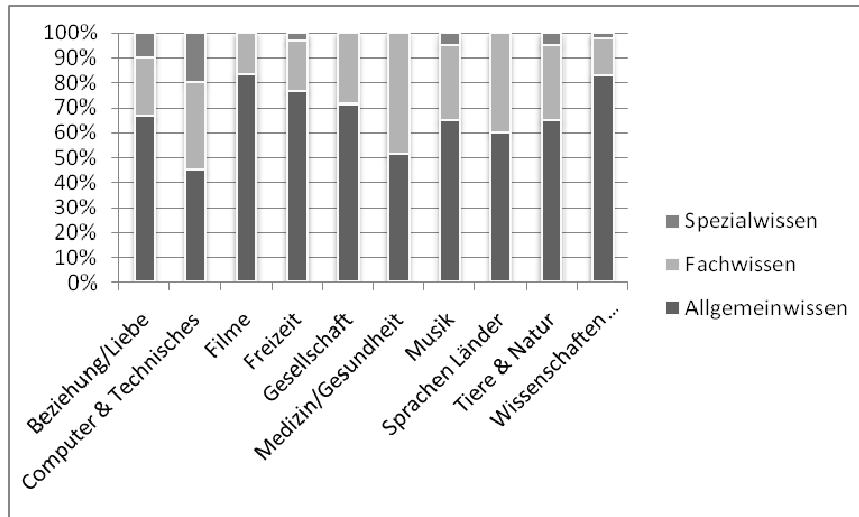


Abb. 5.12 Schwierigkeitsgrad der Fragen in Abhängigkeit vom Themengebiet

Es wurde analysiert, welches Wissen der Antwortende zur Beantwortung der Frage benötigt. Hier zeigt sich ein Zusammenhang zwischen dem Anteil der Fragen, die man als Anfänger beantworten kann, und der oben aufgezeigten Antwortqualität pro Themengebiet. Je höher der Anteil an Allgemeinwissensfragen, desto besser die Antwortqualität.

Der Anteil der Fragen, für die Spezialwissen benötigt wird, ist insgesamt extrem gering. Eine Ausnahme bildet hier der Bereich „Medizin/Gesundheit“, was aber nicht sonderlich überrascht. Interessant ist, dass im Bereich „Filme“, dem Bereich mit den besten Antworten, die leichtesten Fragen gestellt werden. Dazu passt, dass alle Fragen, die man als Anfänger beantworten kann, im Schnitt eine Antwort-Punktzahl von 73,03 erreichen, während die Fragen für Fortgeschrittene nur eine Antwortqualität von 66,24 und die Fragen für Spezialisten sogar nur eine Antwortqualität von 62,35 erreichen. Der Schwierigkeitsgrad der Frage scheint also direkten Einfluss auf die Qualität der Antworten zu haben, auch wenn die Unterschiede weniger gravierend ausfallen, als das eventuell zu erwarten gewesen wäre.

Dies erklärt die Ausnahmestellung des Bereiches „Filme“ bei der Antwortqualität in Teilen. Da allerdings auch der Bereich „Wissenschaften“ nur relativ einfache Fragen aufweist, müssen neben der Schwierigkeit der Frage auch andere Faktoren wie zum Beispiel das Interesse am Thema der Frage eine Rolle spielen.

Stellt man den Schwierigkeitsgrad in Abhängigkeit von der Fragen-Art dar, so lässt sich die oben gewonnene Erkenntnis nicht bestätigen (vgl. Abb. 5.13). Die Verteilung zeigt zwar, dass die qualitativ eher schlechter beantworteten Meinungsfragen auch den höchsten Schwierigkeitsgrad aufweisen, die qualitativ deutlich besser beantworteten Faktenfragen weisen jedoch einen nahezu ebenso hohen Schwierigkeitsgrad auf. Die Spaßfragen mit dem logischerweise niedrigsten Schwierigkeitsgrad generieren sogar die schlechtesten Antworten.

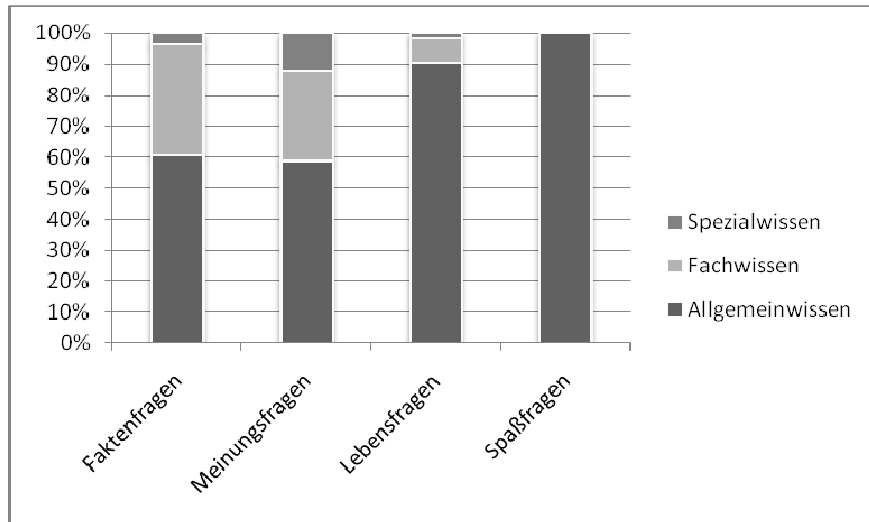


Abb. 5.13 Schwierigkeitsgrad der Fragen in Abhängigkeit von der Art der Frage

Setzt man den Schwierigkeitsgrad in direkte Abhängigkeit zur Qualität der Antworten, so zeigt sich, dass gute Antworten zwar tendenziell einen etwas höheren Anteil an einfachen Fragen aufweisen, der Zusammenhang aber recht gering ist (vgl. Abb. 5.14). Die Qualität der Antworten lässt sich also nicht allein durch den Schwierigkeitsgrad der Frage beeinflussen. Je nach Art oder Thema der Frage variiert die Antwortqualität teils erheblich.

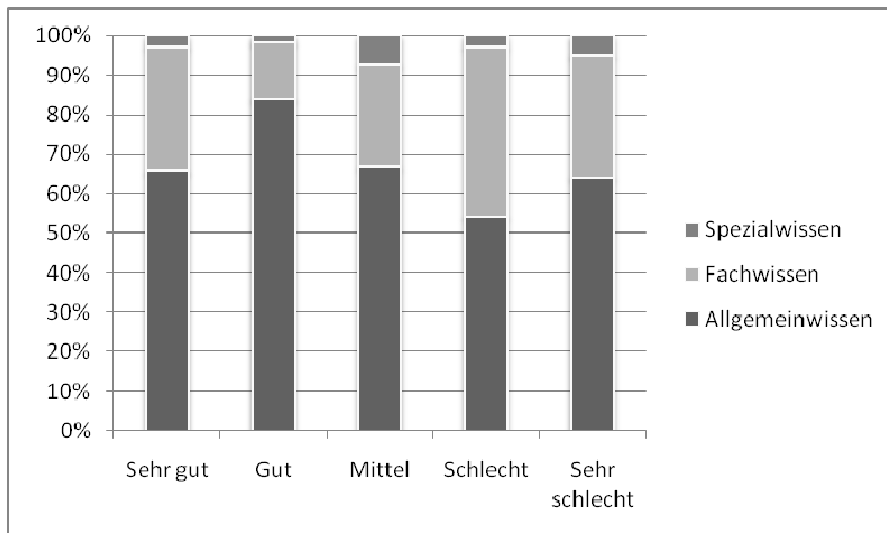


Abb. 5.14 Schwierigkeitsgrad der Fragen in Abhängigkeit von der Antwortqualität

8.3 Anzahl der Antworten

Ein weiteres analysiertes Merkmal, das zwar nicht in die Berechnung der Antwortqualität eingeflossen ist, aber durchaus von Bedeutung sein könnte, ist die Anzahl der Antworten, die pro Frage gegeben wurden. Stellt man die Anzahl der Antworten und die Antwortqualität für die einzelnen Themen dar (vgl. Abb. 5.15), so sieht man, dass sich die Anzahl der Antworten eher antiproportional zur Antwortqualität verhält. Dieser Zusammenhang ist zwar sehr schwach ausgeprägt, aber trotzdem lässt sich festhalten, dass die Qualität der Antworten bei Fragen mit vielen Antworten tendenziell schlechter ist als bei Fragen, auf die nur wenig User antworten. Ein Vergleich mit Abbildung 5.14 zeigt, dass der Schwierigkeitsgrad der Frage keinen direkten Einfluss auf die Anzahl der Antworten hat.

Innerhalb der Themen generieren Fragen aus dem Themenbereich „Sprachen/Länder“ die meisten Antworten pro Frage. Die geringste Anzahl Antworten pro Frage wird im Themenbereich „Computer & Technisches“ generiert.

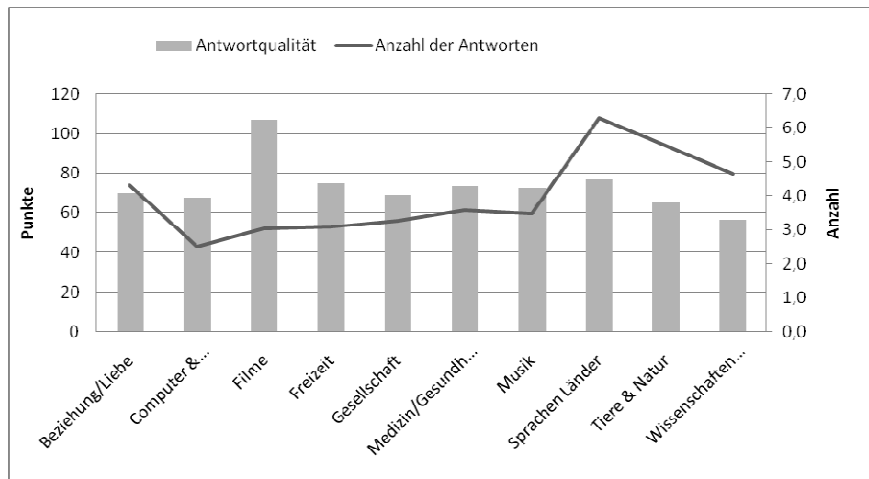


Abb. 5.15 Durchschnittliche Antwortqualität und durchschnittliche Anzahl der Antworten pro Frage in den Themengebieten.

Interessant ist, dass die Anzahl der Antworten nicht nur vom Thema, sondern auch von der Art der Frage abzuhängen scheint. Faktenfragen, die wie bereits gezeigt, die beste Antwortqualität generieren und auch insgesamt den größten Anteil der Fragen ausmachen, generieren am wenigsten Antworten pro Frage (vgl. Abb. 5.16). Die aufgestellte Hypothese, bei Faktenfragen, also bei Fragen mit meist nur einer Antwortmöglichkeit, könnte die Anzahl der Antworten geringer sein, scheint sich hier zu bewahrheiten. Grund hierfür könnte sein, dass sich bei nur einer Antwortmöglichkeit schnell Dopplungen in den Antworten ergeben und die User daher keinen Anreiz für weitere Antworten sehen.

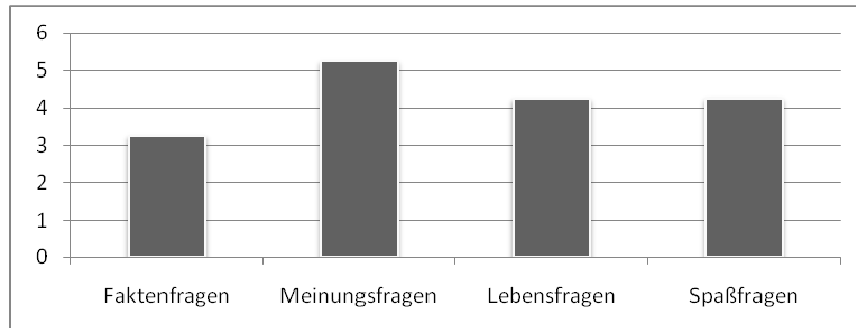


Abb.5.16 Anzahl der Antworten pro Frage-Art

8.4 Bonuspunkte

Stellt man die Anzahl der durchschnittlich ausgesetzten Bonuspunkte pro Frage und die durchschnittliche Antwortqualität für die verschiedenen Themen dar, so zeigt sich, dass die Themen mit hoher Antwortqualität im Schnitt weniger Bonuspunkte pro Frage aufweisen. Die Anzahl der Bonuspunkte verhält sich dabei teilweise sogar antiproportional zur Antwortqualität. Dies zeigt sich vor allem im Themenbereich „Filme“ und wiederum im Themenbereich „Computer & Technisches“ (vgl. Abb. 5.17).

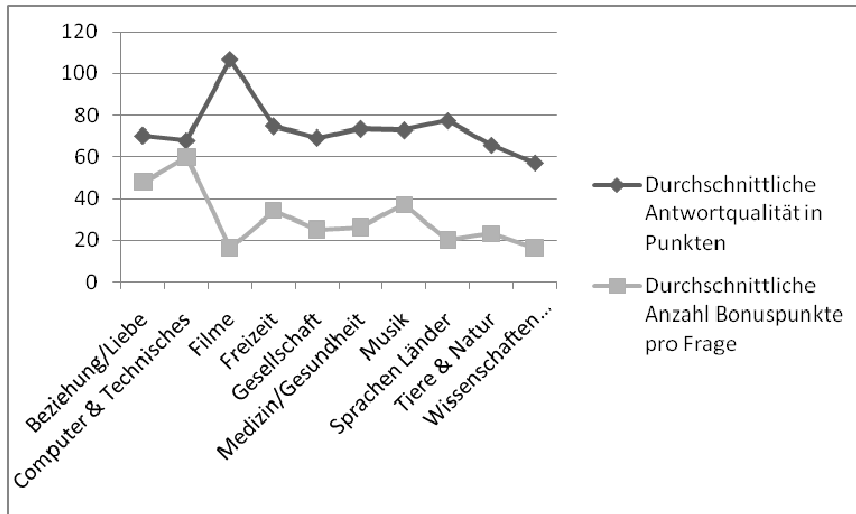


Abb. 5.17 Bonuspunkte für Fragen in Abhängigkeit zur Antwortqualität in den Themenbereichen

Die Anzahl der ausgesetzten Bonuspunkte scheint erstaunlicherweise also keine bzw. negative Auswirkungen auf die Antwortqualität zu haben. Eine mögliche Erklärung hierfür wäre, dass viele Bonuspunkte zu unüberlegten Antworten einladen, mit denen versucht wird, trotz mangelnder Sachkenntnis Punkte für das eigene Konto

zu sammeln. Sollte diese Hypothese zutreffen, müsste sich aber eine Korrelation zwischen der Anzahl der Antworten pro Frage und der Höhe der Bonuspunkte ergeben. Abbildung 5.18 zeigt, dass eine solche Korrelation nicht besteht. Eine andere Erklärung wäre, dass Fragesteller für schwierige Antworten tendenziell mehr Bonuspunkte aussetzen, diese Fragen von den Usern im Durchschnitt aber schlechter beantwortet werden. Hierauf würde auch der deutlich antiproportionale Zusammenhang zwischen Bonuspunkten und Antwortqualität bei Fragen aus dem Bereich „Filme“ hindeuten. Die negative Auswirkung der Bonuspunkte auf die Antwortqualität, bzw. der Zusammenhang zwischen Bonuspunkten und Antwortqualität wäre aber noch näher zu untersuchen.

Deutlicher ist der Zusammenhang zwischen der Anzahl der Bonuspunkte und der Anzahl der positiven Bewertungen, die die Antworten zu einer Frage insgesamt bekommen. Hier führen hohe Bonuspunkte offensichtlich zu vielen Gut-Bewertungen. Die extrem hohe Anzahl der Bonuspunkte im Bereich der Spaßfragen, die für den nicht ernst gemeinten Charakter dieser Fragen spricht, kann den Wert der positiven Bewertungen allerdings nicht in gleichem Maße steigern. Warum dieser Zusammenhang zwischen der Anzahl der Gut-Bewertungen und der Höhe der Bonuspunkte besteht, lässt sich nur spekulieren. Möglich wäre, dass User, die viele Bonuspunkte aussetzen, die erhaltenen Antworten einfach auch aktiver bewerten als User, die auf das Aussetzen von Bonuspunkten weniger Wert legen.

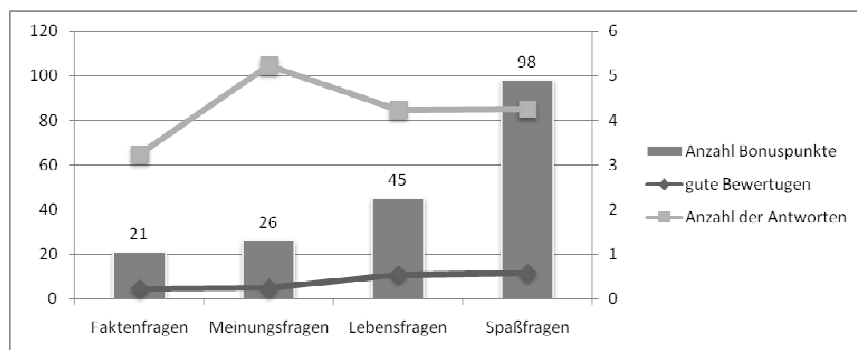


Abb. 5.18 Korrelationen: Bonuspunkte, Gut-Bewertungen, Antwortenanzahl

8.5 Gut-Bewertungen

Neben dem Zusammenhang zwischen Gut-Bewertungen und der Anzahl der Bonuspunkte ist auch der Zusammenhang zwischen der Anzahl der Gut-Bewertungen und der Antwortqualität interessant. Mit Ausnahme der guten Antworten ist die Anzahl der Gut-Bewertungen bei allen Qualitätsstufen recht ähnlich, wenn auch bei sehr schlechten Fragen erwartungsgemäß am niedrigsten (vgl. Abb. 5.18). Trotzdem scheint auch die Bewertung der Antworten durch die User kein geeignetes Kriterium zur Bestimmung guter Antworten zu sein. Dieser Eindruck wird durch Abbildung 5.19 noch verstärkt. Hier zeigt sich, dass sich die Anzahl der guten Bewertungen in Abhängigkeit von der Art der Frage eher antiproportional zur tatsächlichen Antwortqualität verhält. Dieses Ergebnis wird dabei natürlich durch den zu Grunde gelegten Kriterienkatalog ermöglicht, der die Qualität der Antworten nicht nur in

Abhängigkeit von der User-Bewertung, sondern vor allem auf Grund einer durch unabhängige Probanden vorgenommenen Bewertung einordnet.

Eine gute Bewertung ist also kein geeignetes Indiz für eine gute Antwort. Im Gegenteil scheint eine gute Bewertung vor allem bei den Faktenfragen und den Meinungsfragen eher auf eine geringere Qualität der Antworten hinzuweisen. Dieser Zusammenhang ist auf Grund der gänzlich anderen Erwartungshaltung besonders überraschend.

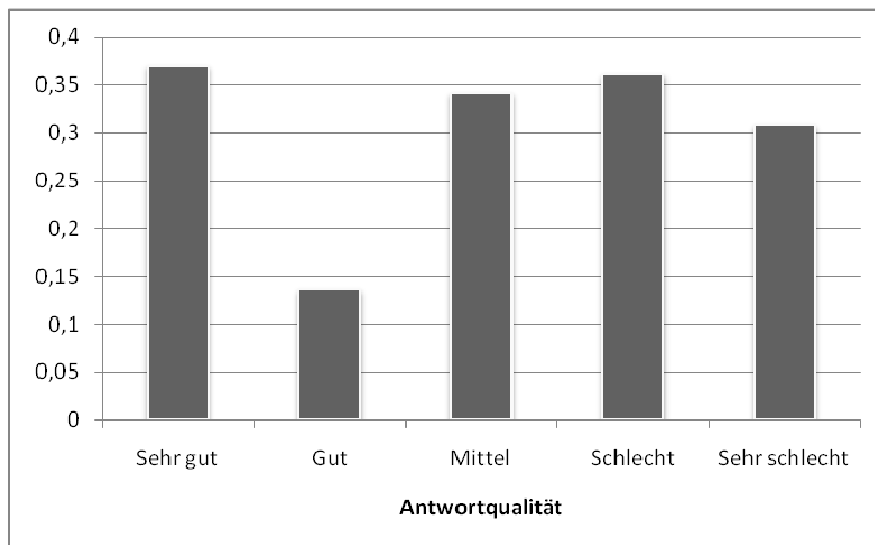


Abb. 5.18 Anzahl der Gut-Bewertungen pro Frage in Abhängigkeit von der Antwortqualität, n=247

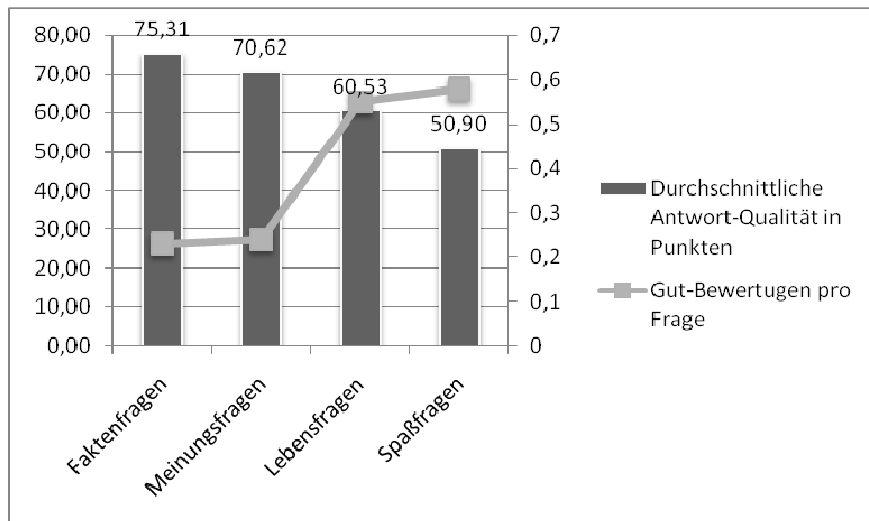


Abb. 5.19 Anzahl der Gut-Bewertungen und Antwort-Qualität in Abhängigkeit von der Art der Frage, n=247

8.6 Formale Merkmale

Stellt man die Qualität der formalen Merkmale in Abhängigkeit von der Antwortqualität dar, so lässt sich kein genereller Zusammenhang feststellen (vgl. Abb. 5.20). Die Abbildung stellt dabei die anteiligen Fragen dar, die für das jeweilige formale Merkmal positiv bewertet wurden.

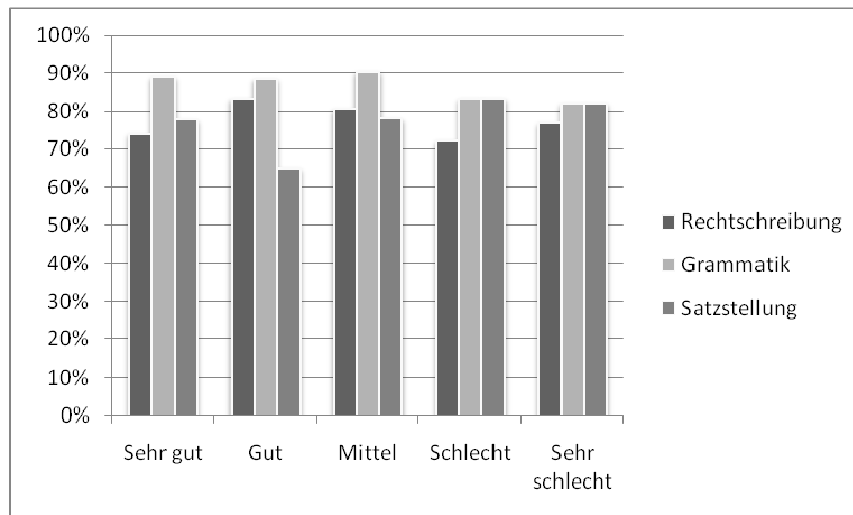


Abb. 5.20 Formale Merkmale in Abhängigkeit zur Antwortqualität

Tendenziell erscheint es zwar so, dass eine gute Grammatik zumindest eine schlechte Antwort unwahrscheinlicher macht, ein Kriterium zur Beurteilung der Qualität einer Frage lässt sich so aber sicherlich nicht ableiten. Auch die Rechtschreibung hat offensichtlich keinen Einfluss auf die Antwortqualität.

Allerdings lässt sich ein Zusammenhang zwischen der formalen Qualität einer Frage als Ganzes und der Antwortqualität in Abhängigkeit vom Thema der Frage zumindest ansatzweise erkennen (vgl. Abb. 5.21). Dies wird vor allem in den Bereichen „Filme“ und „Sprachen/Länder“ deutlich. Hier wirken sich gute formale Merkmale also positiv auf die Antwortqualität aus.

Der insgesamt geringe Einfluss der formalen Gestaltung der Frage auf die Qualität der Antworten ist bei genauerer Überlegung aber auch nicht weiter verwunderlich. Es ist davon auszugehen, dass ein User, wenn er die Frage verstanden hat, bei einer formal schlecht gestalteten Frage genauso „klug“ und richtig antworten dürfte, wie bei einer formal sehr gut gestalteten Frage.

Weiterhin ist festzustellen, dass die formalen Merkmale bei Fragen aus dem Bereich „Sprachen/Länder“ deutlich besser sind als zum Beispiel im Bereich „Computer & Technisches“.

Interessant ist auch der relativ deutliche Zusammenhang zwischen der Qualität der formalen Merkmale und der Anzahl der Antworten. Offensichtlich generieren gut gestaltete Fragen mehr Antworten als Fragen mit formalen Mängeln. Dies impliziert zwar keine Aussage über die Qualität, da eine hohe Anzahl von Antworten aber

sicherlich recht positiv auf den fragenden User wirkt, ist dieses Kriterium für ein späteres Tutorial (vgl. Kap. 10) von Bedeutung.

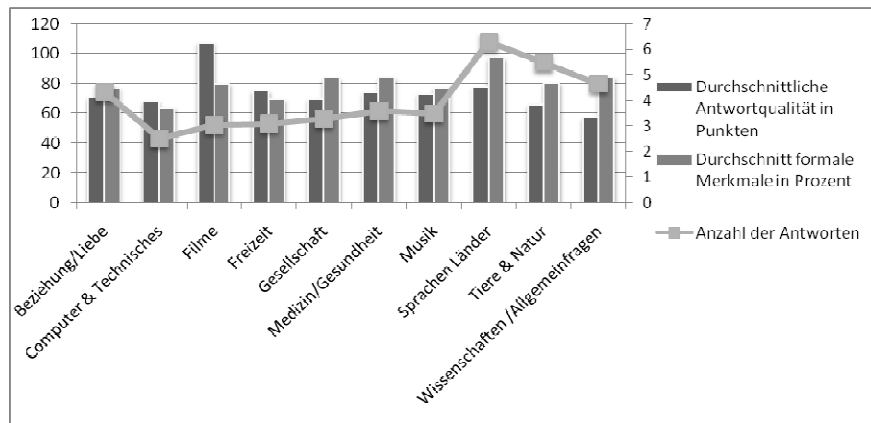


Abb. 5.21 Formale Merkmale der Frage und deren Antwortqualität in Abhängigkeit vom Themengebiet

Positiv anzumerken ist insgesamt die durchgehend hohe formale Qualität der Fragen, die in einer Online-Community, denkt man an den Sprachstil in vielen Internet-Foren, so nicht zu erwarten war.

8.7 Inhaltliche Merkmale

Welche inhaltlichen Kriterien sind wesentlich für eine gute Frage? Diese Frage versuchen wir mit der folgenden Auswertung zu beantworten (vgl. Abb. 5.22). Einige Rückschlüsse lassen sich hier auf den ersten Blick feststellen. Zum einen senkt der Verzicht auf eine provokative Wortwahl die Gefahr einer schlechten Antwort deutlich, zum anderen sind fast alle Fragen mit sehr guten Antworten in sich logisch und ein Großteil der Fragen mit schlechter Antwortqualität lassen eben diese Logik vermissen.

Der Anteil der Fragen mit mehr als einer Antwort-Möglichkeit ist bei den sehr schlechten Antworten deutlich höher als bei den übrigen Qualitätsstufen.

Die anderen untersuchten Kriterien scheinen in keinem direkten Zusammenhang mit der generierten Antwortqualität zu stehen. Erstaunlicherweise unterscheiden sich weder der Anteil der Fremd- und Fachwörter noch der Anteil der zeitlich begrenzten Fragen bei Antworten unterschiedlicher Qualität. Wie auch schon bei den formalen Kriterien scheint die Antwortqualität von der Gestaltung der Frage kaum abzuhängen. Nur inhaltliche Merkmale, die das Verstehen der Frage erschweren oder zu einer schlecht bewerteten Antwort provozieren, wirken sich auch auf die Qualität der Antworten aus.

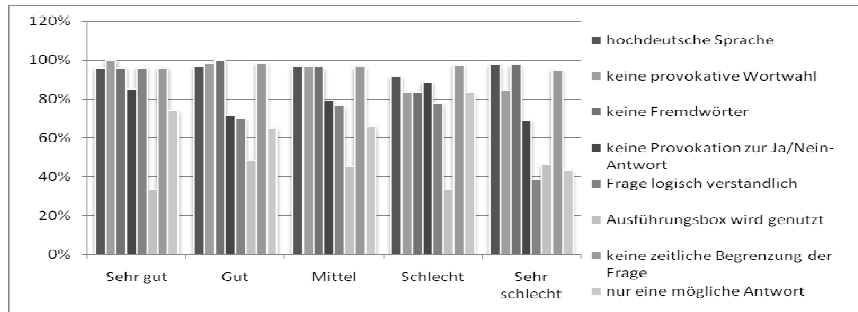


Abb. 5.22 Anteil der Fragen in Abhängigkeit zur Antwortqualität, die das jeweilige Kriterium erfüllt

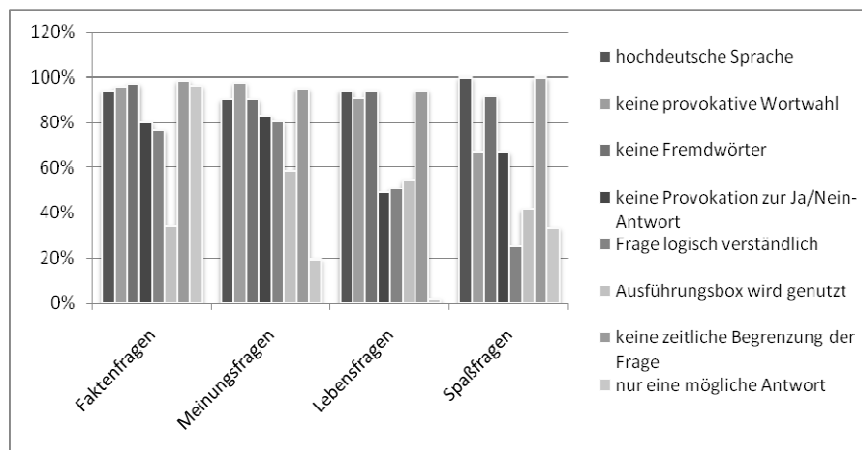


Abb 5.23 Anteil der Fragen in Abhängigkeit zur Art der Frage, die das jeweilige Kriterium erfüllt

Ordnet man die Ausprägung der unterschiedlichen Kriterien den verschiedenen Fragearten zu, so bestätigt sich erneut die hohe Qualität der Faktenfragen (vgl. Abb. 5.23). Auffällig ist hier auch noch einmal, dass fast ausschließlich Faktenfragen nur eine Antwort zulassen. Zu dieser Bestimmtheit der Frageart passt auch, dass die Ausführungsbox bei den Faktenfragen am wenigsten gebraucht wird. Auch der geringe Anteil von Fragen ohne provokative Wortwahl bei den Spaßfragen passt genau ins Bild/erfüllt die Erwartungen (aber warum?).

9. EINBINDEN VON FRAGEN AUS LYCOS IQ IN ANDERE SUCHFORMEN

Anfänglich haben wir die Frage gestellt, inwieweit sich Fragen aus Lycos iQ zur Ergänzung anderer Suchformen eignen. Nach Abschluss der Untersuchung ist festzustellen, dass sich Spaßfragen und eventuell auch Meinungsfragen auf Grund der niedrigeren Qualität nicht für eine Einbindung in andere Suchformen eignen.

Faktenfragen und Lebensfragen sind zwar von leicht unterschiedlicher Qualität, weisen aber insgesamt eine höhere Qualität auf. Generell stellt sich aber die Frage, wie man die Frageart automatisch ermitteln soll, so dass dies nicht einziges Unterscheidungsmerkmal sein kann. Hier sind technische Grenzen gesetzt, die eine komplette Unterscheidung nach der Frageart verhindern. Trotzdem ist es sinnvoll, bevorzugt auf Faktenfragen zurück zu greifen. Eine Möglichkeit diese zu identifizieren wäre zum Beispiel das Anlegen von Wortlisten. Typisch für Faktenfragen sind zum Beispiel Fragewörter wie „Wo“, „Wann“ und „Wer“, während „Wie“ eher bei Lebensfragen auftauchen dürfte. Mit entsprechender Analyse könnte man eine solche Liste erweitern und die Fragen einigermaßen nach ihrer Art filtern.

Hier stellt sich aber auch die Frage, wie interessant Faktenfragen als Bereicherung einer herkömmlichen Suche sind, da sich diese meist zufriedenstellend über algorithmische Suchen beantworten lassen. Interessanter für den User wäre daher vielleicht gerade eine Auswahl der kontrovers diskutierten Fragen und der persönlichen Meinungen, um eine andere Art der Fragen-Beantwortung anzubieten. Hier müsste man aber eine etwas geringere Qualität der Fragen in Kauf nehmen. Dies stellt eine Grundsatzentscheidung dar, die von uns so nicht beantwortet werden kann.

Ein weiteres Kriterium ist das Thema der Frage. Wir haben gezeigt, dass die Antwortqualität in Themen von allgemeinem, freizeitlichem Interesse am besten ist. Dementsprechend sollten bevorzugt Fragen aus diesen Bereichen verwendet werden, während Bereiche, die ein spezielles Fachwissen erfordern, ausgeklammert werden sollten. Passend dazu haben wir gezeigt, dass Fragen, die mit normalem Allgemeinwissen zu beantworten sind, tendenziell die etwas besseren Antworten generieren. Da dies aber zwischen den einzelnen Fragearten und Themen variiert, lässt sich hieraus kein wirklich geeignetes Kriterium ableiten.

In dem Umstand, dass Fragen mit geringem benötigten Fachwissen die deutlich besseren Antworten generieren, liegt allerdings auch ein Problem, da die Lycos iQ-Antworten dann nicht zur Lösung komplizierterer Fragen beitragen können.

Inhaltlich kann gesagt werden, dass geeignete Fragen keine provokative Wortwahl enthalten und zudem logisch aufgebaut sein sollten. Vor allem die Wortwahl könnte anhand weiterer Stoppwortlisten, die die Fragen nach beleidigenden Ausdrücken oder bestimmten Redewendungen durchsuchen, überprüft werden.

Weitere Kriterien, die eine Frage disqualifizieren, sind eher Merkmale, die sich direkt auf der Lycos iQ-Plattform ablesen lassen und die sich daher gut zu einer automatischen Selektion der Fragen eignen. Ein Indiz für qualitativ schlechte Fragen sind so interessanter Weise nach unserer Untersuchung eine hohe Anzahl positiver Bewertungen und eine hohe Anzahl ausgesetzter Bonuspunkte, wobei die Anzahl der positiven Bewertungen in Teilen von der Anzahl der Bonuspunkte abhängt. Die Korrelation zwischen der Anzahl der ausgesetzten Bonuspunkte und der Qualität der Antworten sowie der Anzahl der positiven Bewertungen und der Qualität der Antworten ist also tendenziell negativ.

Zusammengefasst kann also gesagt werden, dass eine Frage, die sich zur Ergänzung einer herkömmlichen Suche eignet, aus einem Themengebiet stammen sollte, das kein spezielles Interesse und Fachwissen erfordert. Außerdem sollten Fragen ausgeklammert werden, die Worte und Formulierungen enthalten, die typisch für Provokationen und Beleidigungen sind. Letztendlich bieten sich eher Fragen mit

einer niedrigen Anzahl positiv bewerteter Antworten und wenigen Bonuspunkten an, da diese tendenziell qualitativ bessere Antworten generieren.

10. TUTORIAL

Bei der Entwicklung eines Tutorials muss man bedenken, dass man einen User beim Einstellen einer Frage eventuell etwas lenken, ihm die Art und das Thema der Frage aber nicht vorschreiben kann. Schließlich soll jede Frage zulässig sein. Einige Kriterien, die die Qualität von Fragen und Antworten verbessern können, bieten sich aber dennoch an, um in Form einer kleinen Hilfe oder eines Tutorials präsentiert zu werden. Folgende Formulierungen wären dabei denkbar:

- Achten Sie auf eine ordentliche formale Gestaltung Ihrer Fragen. Formulieren Sie die Frage im ganzen Satz und achten Sie auf die Rechtschreibung. So erhöhen Sie die Wahrscheinlichkeit, dass Sie viele hilfreiche Antworten auf Ihre Frage erhalten.
- Verzichten Sie auf jegliche provokative Wortwahl oder einen provokativen Charakter ihrer Frage. Beleidigungen oder Angriffe auf Personen oder Gruppen sind nicht nur unerwünscht, sondern führen auch eher zu schlechten Antworten.
- Formulieren Sie Ihre Frage logisch. Fragen ohne innere Logik werden deutlich öfter sehr schlecht beantwortet als logisch formulierte Fragen.
- Achten Sie darauf, dass der Sinn Ihrer Frage den anderen Mitgliedern der Community sofort klar wird.
- Seien Sie sparsam mit Bonuspunkten. Auch bei besonders schwierigen Fragen führt eine hohe Anzahl Bonuspunkte nicht zu einer besseren Antwort.

Jeder Tipp sollte außerdem durch ein entsprechendes, aussagekräftiges Beispiel veranschaulicht werden, an dem der Fragesteller sich orientieren kann. Ideal wäre hier die Anlage eines Tutorials als kleine Videosequenz.

11. FAZIT

Im Rahmen der Analyse konnten wir allgemeine Zusammenhänge zwischen Fragen und Antworten aufzeigen. Einige Korrelationen waren dabei recht deutlich, andere eher schwach und einige der erhofften Zusammenhänge konnten so auch nicht belegt werden. Vor allem Zusammenhänge zwischen inhaltlichen und formalen Kriterien und der Antwortqualität ließen sich nur schwer aufzeigen. Überraschend ist, dass sich die formalen Kriterien zwar insgesamt schwach auf die Qualität der Antworten auswirken, schlechte Formalien die Lycos iQ-User aber nicht wirklich vom Beantworten einer Frage abzuhalten scheinen. Dafür entspricht das in allen Bereichen gute Abschneiden der Faktenfragen unseren Erwartungen.

Ein interessanter Ansatzpunkt für weitere Untersuchungen bietet die Frage, warum die Antwortqualität bei Fragen, die kaum Fachwissen benötigen, am besten ist. Das populäre Schlagwort „Wisdom of the crowd“ im Gegenteil zu dem Wissen einzelner Experten scheint hier eine Rolle zu spielen. Herauszufinden wäre also, ob einfache Fragen aus Themen, die viele Menschen interessant finden, von einer Menge

besser beantwortet werden können als von einzelnen Experten und ob umgekehrt einzelne Experten mit spezielleren Fragen besser umgehen können als die Masse.

Durch eine deutliche Erhöhung des Stichprobenumfangs und eine noch feinere Aufspaltung der einzelnen Kriterien und Themenkategorien könnte eine weiterführende Untersuchung einige Ergebnisse sicherlich noch vertiefend darstellen. Diese Analyse ist daher eher als erster Überblick und als Anreiz für tiefergehende, speziellere Untersuchungen zu verstehen. Trotzdem hoffen wir, dass die Ergebnisse sich bei zukünftigen Entscheidungen und Verbesserungen an der Lycos iQ-Plattform als nützlich erweisen könnten.

Verbesserung des Taggings in Lycos iQ

STEFAN SEEHAUS

stephan.seehaus@haw-hamburg.de

FABIAN LEMM

fabian.lemm@haw-hamburg.de

Um Möglichkeiten zu finden, das Tagging in Lycos iQ zu verbessern, wurden verschiedene Anbieter untersucht, die das freie Tagging ebenfalls in ihr Angebot integriert haben. So wird empfohlen z. B. ein Videotutorial, Kategorien oder erweiterte Bearbeitungsmöglichkeiten für Tags einzuführen, da diese Features? das Tagging verbessern können. Darüber hinaus wird u. a. vorgeschlagen, die Tagging-Rechte zu erweitern, damit auch Personen, die auf eine Frage reagieren, ihre Antwort taggen können. Die Antwort bzw. die Frage wird dadurch mit zusätzlichen Tags ergänzt, so dass diese hier als relevante Suchbegriffe mit eingeführt und zudem Rechtsschreibfehler korrigiert werden können.

1. EINLEITUNG

Im Rahmen des Projektes war es das Ziel, das Tagging von Lycos iQ zu analysieren und Vorschläge für mögliche Verbesserungen zu machen. Hierzu wurden verschiedene Anbieter analysiert, die das Tagging in ihre Angebote mit eingebunden haben, um so einen Vergleich herzustellen. Im folgenden Kapitel wird näher auf die Probleme eingegangen, die beim Tagging von Lycos iQ festgestellt wurden. Daraufhin werden die Kriterien erklärt, nach denen die anderen Dienste untersucht wurden und die jeweiligen Ergebnisse werden vorgestellt. Abschließend wird eine Empfehlung zur Optimierung des Taggings bei Lycos iQ abgegeben. Eine Verbesserung des Taggings hätte zur Folge, dass die Fragen besser zu finden wären. Dies kann bei der Implementierung von Fragen bzw. Antworten in klassischen Suchmaschinen helfen, denn mit einer besseren Qualität des Taggings könnten die Fragen, die angezeigt werden, auch eher passend zum in der Suchmaschine eingegebenen Thema oder Stichwort sein.

2. PROBLEME BEIM TAGGING

Mit der freien Verschlagwortung durch den Nutzer gehen zahlreiche Probleme einher. Wird z.B. eine Frage zum Thema „Auto“ getaggt, wählen die unterschiedlichen Nutzer unterschiedliche Tags: Neben „Auto“ werden beispielsweise auch „Autos“, „Automobil“, „Automobile“, „KFZ“, „Wagen“ oder eine direkte Typ-Bezeichnung wie „Audi A8“ gewählt. Da praktisch jedes Wort in den beschriebenen Systemen als Tag gewählt werden kann, entstehen so zwangsläufig folgende Probleme:

2.1 Synonyme

Von Synonymen spricht man, wenn zwei Wörter die gleiche Bedeutung haben. Benutzt ein User für seine Frage das Tag „Internet“, bevorzugt ein anderer für die gleiche Quelle „Netz“ oder „Web“. Bei einer Suche mit nur einem der Begriffe entgehen dem Nutzer also unter Umständen für ihn relevante Treffer. Die Treffermengen sind bei jedem der Begriffe unterschiedlich, da die User jeweils unterschiedliche Begriffe bzw. Synonyme verwendet haben.

Die Anzeige von verwandten Themen bei Lycos iQ hilft aber dabei, verwandte Suchbegriffe mit in die Suche einzubeziehen, da sich unter diesen Begriffen auch Synonyme befinden können. Bei einer Suche nach „Computer“ werden auch „PC“ oder „Software“ angezeigt. Bei Themen, die weniger populär sind, fallen die Ergebnisse allerdings schlechter aus.

2.2 Homonyme

Hat ein Wort mehrere Bedeutungen ist dies ebenfalls problematisch. Das Wort „Kette“ kann verschiedene Bedeutungen haben. Dieses Tag findet sich bei Fragen zu völlig verschiedenen Themen wie Schmuck, einer Handelskette oder einer Fahrradkette.

Mit der richtigen Nutzung der Suche lässt sich diese Schwierigkeit bei Lycos iQ umgehen oder zumindest eindämmen: Da eine kombinierte Suche mit mehreren Tags möglich ist, kann man weitere Suchbegriffe mit einbeziehen und so leichter Treffer in dem gewünschten Kontext erzielen.

2.3 Abkürzungen / Akronyme

Ein Akronym ist ein Kunstwort, das aus den Anfangsbuchstaben mehrerer Wörter zusammengesetzt wird. Abkürzungen und Akronyme werden ebenfalls als Tag verwendet: „KFZ“ statt „Kraftfahrzeug“ oder „SEO“ statt Suchmaschinenoptimierung. Die Problematik ist hier die gleiche wie bei den Synonymen, da es auch hier zu einer Zersplitterung kommt.

2.4 Singular- und Pluralform

Während bei der Verschlagwortung mithilfe von kontrollierten Vokabularen in der Regel die Singularform eines Schlagworts verwendet wird, ist es beim freien Tagging dem User überlassen, welche Form er wählt. Wenn nun die eine Hälfte der User das Tag „Buch“ wählt und die andere Hälfte „Bücher“, gibt es auch hier wieder verschiedene Tags für denselben Begriff.

Durch die Verwendung des suggestiven Taggings, bei dem während der Eingabe bereits Tags vorgeschlagen werden, lassen sich solche Probleme zumindest beschränken.

2.5 Rechtschreibung

Unterläuft bei der Eingabe des Tags ein Rechtschreibfehler oder Tippfehler, kann man davon ausgehen, dass die Ressource nur schwer wieder gefunden werden kann. Zudem werden verschiedene Schreibweisen verwendet, wie z.B. „Photo/ Foto“.

2.6 Zusammengesetzte Begriffe

„Open Source“ oder „OpenSource“? Beide Schreibweisen sind möglich. Del.icio.us empfiehlt in seiner Hilfe dem User z.B. jede Schreibweise zu nutzen, die ihm richtig erscheint. Ob „SanFrancisco“, „san.francisco“, „san-francisco“ oder einfach „sf“, entscheidet der User [1]. Auch hier führen die vielen Varianten zu einer Zersplitterung.

2.7 Spezifizierung

Tags wie „Allgemein“ oder „Fragen“ sind nicht aussagefähig, da sie zu unspezifisch sind und keinen Rückschluss auf den Inhalt zulassen. Auch der umgekehrte Fall ist möglich: Wenn jemand eine sehr spezifische Beschreibung wählt, da er sich beruflich oder privat sehr intensiv mit einem Thema beschäftigt, so ist das verwendete Vokabular wesentlich spezieller als bei anderen Usern. Es wird z.B. „Glatthaar-Foxterrier“ als Tag gewählt, während jemand anderes sich einfach für das Tag „Hund“ entscheidet.

2.8 Sprache

Da Lycos iQ Deutschland ein Dienst für deutschsprachige User ist, fällt die Sprachbarriere hier nicht so sehr ins Gewicht wie beispielsweise bei flickr, wo User aus der ganzen Welt ihren Fotobestand pflegen und teilen. Dennoch ist es denkbar, eine Frage zu einem gesuchten Lied auch mit „Song“ zu taggen. Die Problematik ist hier wieder einmal ähnlich wie bei den Synonymen: Für ein und dieselbe Sache können unterschiedliche Tags vergeben werden.

Trotz dieser Probleme beim Tagging stellt die Vergabe von Schlagworten und das Teilen dieser Tags zwischen Benutzern eine gute Möglichkeit zur gemeinsamen Klassifizierung von Informationsräumen dar. Gleichzeitig entstehen dadurch auch bessere Suchmöglichkeiten. Durch die einfache Nutzbarkeit und den geringen Aufwand erreicht man eine hohe Beteiligung und damit auch Netzwerkeffekte. [2]

Im Gegensatz zu kontrollierten Vokabularen und Klassifikationssystemen ist der Einstieg in das freie Tagging wesentlich einfacher für den User, da das Konzept leichter zu verstehen ist. Vorteilhaft am freien Tagging ist zudem die sprachliche Flexibilität, die die Sprache der Nutzer ernst nimmt und nicht normierend eingreift. Die User können also immer die Tags wählen, die ihnen am passendsten erscheinen. Auch wenn dies zu einer großen Vielfalt führt, ließe sich das Funktionieren des Taggings mit einer Selbstregulation durch die Gesamtheit der User erklären, da trotz der großen Unterschiede bei den Usern gewisse Tags für einen Inhalt besonders häufig vergeben werden. [3]

Zusätzlich können die oben angesprochenen Schwierigkeiten teilweise durch Funktionalität der Benutzungsschnittstelle (z.B. Anzeige aller bisher verwendeten oder passenden Tags) oder die zusätzliche Anzeige verwandter Begriffe bei Suchergebnissen abgemildert werden.

3. KRITERIEN FÜR DIE ANALYSE DER UNTERSCHIEDLICHEN DIENSTE

Um die verschiedenen Dienste analysieren zu können, wurden die folgenden Kriterien festgelegt (vgl. Abb. 6.1). Die Kriterien „Tagging Support“ und „Tagging Rights“ sind dem Artikel über den Vergleich des Taggings von C. Marlow et al [3] entnommen.

Anbieter	Tagging	Tagging Support	Verwandte Tags enthalten in Suchergebnissen	Besonderheiten
Lycos iQ iq.lycos.de	ja	suggestive	ja	Tags lassen sich bei der Suche kombinieren
fragr.de www.fragr.de	ja	suggestive	nein	zusätzlich Kategorien
Yahoo! Clever de.answers.yahoo.com	nein	x	x	nur Kategorien, kein Tagging
wer-weiss-was www.wer-weiss-was.de	nein	x	x	kein Tagging
willwissen.net www.willwissen.net	nein	x	x	nur Kategorien, kein Tagging
Quansr quansr.itopia.de	ja	blind	nein	kein Suchfeld
gutefrage.net www.gutefrage.net	ja	suggestive / viewable	nein	keine kombinierte Suche
motorradfrage.net www.motorradfrage.net	ja	suggestive	nein	thematisch stark eingegrenzt
sportlerfrage.net www.sportlerfrage.net	ja	suggestive	nein	thematisch stark eingegrenzt

Abb. 6.1 Übersicht der Analysekriterien

Ein Beispiel für „viewable“ und „suggestive tagging“ findet sich auf Abb. 6.2. Die Vorteile von „suggestive“ und „viewable tagging“ sind, dass durch die vorgeschlagenen bzw. angezeigten Tags eine größere Einheitlichkeit der Verschlagwortung erreicht wird. Die somit verbesserte Qualität der Tags erleichtert wiederum die Suche der Nutzer nach Einträgen. .

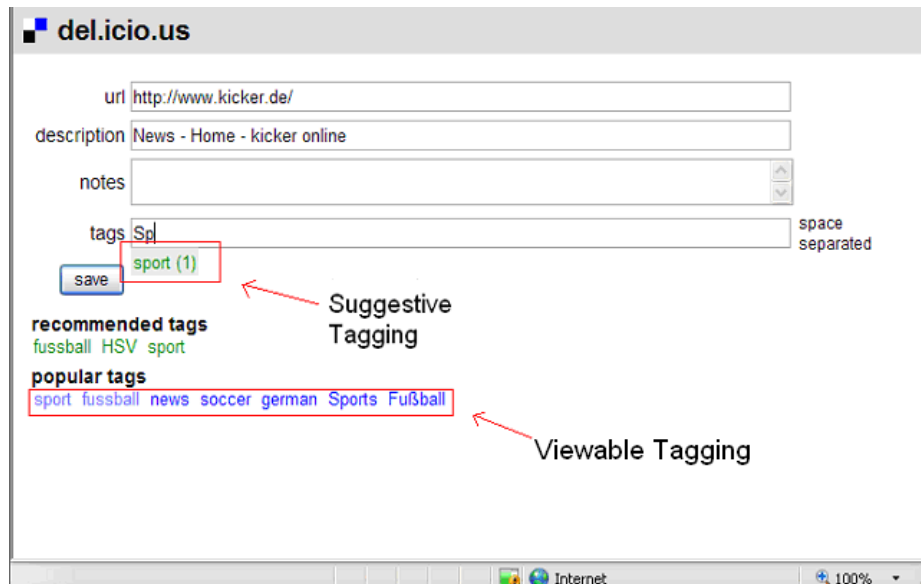


Abb. 6.2 Eingabemaske von del.icio.us

4. ANALYSIERTE DIENSTE

Um eine gute Übersicht über den Status quo des Taggings zu bekommen, aber auch um Anregungen für die Verbesserung der Tagging-Situation bei Lycos iQ zu finden, wurden eine Reihe von Diensten untersucht, die das Tagging in ihr Angebot ebenfalls integriert haben. Viele Angebote zeigen Gemeinsamkeiten, besonders wenn sie zur selben Art von Dienst gehören. Sie wurden für diese Untersuchung in drei Gruppen eingeteilt:

- Frage-Antwort-Dienste
- Social Bookmarking-Dienste
- Andere Web-Anwendungen

In der Gruppe „andere Web-Anwendungen“ sind verschiedene Dienste zusammengefasst wie z.B. ein Online-Versandhandel, Foto- und Videoportale, ein Musikportal, ein Bibliothekskatalog und ein Social Networking-Portal. Es muss darauf hingewiesen werden, dass die Ergebnisse durch Weiterentwicklung der einzelnen Dienste und Web-Anwendungen veraltet sein könnten. Der Frage-Antwort-Dienst von Lycos iQ wird in Abb. 6.3 und zusätzlich dazu der Linkkatalog von Lycos iQ in Tabelle 6.6 als Social Bookmarking-Dienst aufgeführt.

4.1 Frage-Antwort-Dienste

Frage-Antwort-Dienste sind Web-Anwendungen, bei denen man, im Gegensatz zu den klassischen Suchmaschinen, Suchanfragen in Form von ganzen Sätzen stellen kann. Die Fragen werden von anderen Nutzern der Webseite beantwortet. Es können so Antworten auf Fragen erhalten werden, die bei einer Suchmaschine nur schwer oder überhaupt nicht gefunden werden können. Die Nutzer, die eine Frage stellen, taggen bei Frage-Antwort-Diensten die gestellten Fragen.

Bei der Auswahl der Dienste wurde eine Mischung aus älteren Anbietern wie Wer Weiss Was und neueren Diensten, die sich noch in Beta-Phasen befinden, wie z.B. Quansr ausgewählt. Außerdem wurden Yahoo! Clever und willwissen.net untersucht, welche ein Kategoriensystem benutzen. Einen Überblick über die einzelnen Frage-Antwort-Dienste bietet Abb. 6.3, in der die folgenden Kriterien dargestellt werden: "Tagging" zeigt an, ob der jeweilige Dienst Tagging anbietet; „Tagging Support“ =?; „Verwandte Tags, enthalten in Suchergebnissen“ zeigt an, ob bei den Ergebnissen einer Suche auch auf verwandte Tags verwiesen wird. Außerdem werden „Besonderheiten“ der jeweiligen Anbieter angegeben, wie z.B. die Möglichkeit bei Lycos iQ Tags mit einander zu kombinieren. Die Spalte „Tagging Rights“ ist hier nicht vertreten, da alle untersuchten Frage-Antwort-Dienste, bei denen Tagging benutzt wird, „self-tagging“ verwenden.

Abb. 6.3 Übersicht Frage- und Antwort-Dienste

Anbieter	Tagging	Tagging Support	Verwandte Tags enthalten in Suchergebnissen	Besonderheiten
Lycos iQ iq.lycos.de	ja	suggestive	ja	Tags lassen sich bei der Suche kombinieren
fragr.de www.fragr.de	ja	suggestive	nein	zusätzlich Kategorien
gutefrage.net www.gutefrage.net	ja	suggestive / viewable	nein	keine kombinierte Suche
sportlerfrage.net www.sportlerfrage.net	ja	suggestive	nein	thematisch stark eingegrenzt
motorradfrage.net www.motorradfrage.net	ja	suggestive	nein	thematisch stark eingegrenzt
Quansr quansr.itopia.de/	ja	blind	nein	kein Suchfeld
Yahoo! Clever de.answers.yahoo.com	nein	x	x	nur Kategorien, kein Tagging
wer-weiss-was www.wer-weiss-was.de	nein	x	x	kein Tagging
willwissen.net www.willwissen.net	nein	x	x	nur Kategorien, kein Tagging

Abb. 6.3 Übersicht Frage- und Antwort-Dienste

4.1.1 fragr.de

Dieser Dienst befindet sich noch in einer Beta-Pphase, ist also noch nicht komplett fertig gestellt. Dies könnte auch das Fehlen jeglicher Hilfetexte zum Thema Tagging erklären. Bei der Eingabe eines Wortes in das Feld „Tags“ gibt es eine „suggestive tagging“-Funktion.

Eine weitere interessante Funktion schlägt automatisch Tags vor, die aus dem eingegebenen Fragetext generiert werden. Man muss allerdings vorher die Frage per Button „Zwischenspeichern“ absichern, damit diese Vorschläge angezeigt werden. Es werden Wörter vorgeschlagen, welche der Nutzer im Titel oder im Text eingegeben hat. Leider sind nicht alle Vorschläge sinnvoll. So werden Wörter und Abkürzungen vorgeschlagen, die in den eingegebenen Wörtern enthalten sind, z.B. „ag“ als Tag für „Frage“ oder „eis“ für „Meister“.

Ebenfalls interessant ist, dass man zusätzlich zu seinen Tags aus 37 Kategorien weitere Tags auswählen kann (vgl. Abb. 6.4). Die Auswahl von Kategorien ist hier nicht Pflicht. Leider sind nicht alle Kategorien gut gewählt worden, beispielsweise ist „Alles und Nichts“ nicht aussagekräftig und kann im Grunde auf alle Fragen zutreffen.

Fragen suchen kann man über ein Suchfeld, aber auch über die Kategorien oder eine Tagcloud. Eine Tagcloud ist eine Liste von Tags, die meistens in alphabetischer Reihenfolge stehen. Die Tags, die häufiger vergeben worden sind, werden in einer Tagcloud größer dargestellt als die Tags, die weniger häufig vergeben wurden. Bei den Kategorien wird angezeigt, wie viele Fragen in der jeweiligen Kategorie bereits existieren.

The screenshot shows the 'fragr.de' input form. At the top, there's a navigation bar with 'Tellerrand', 'Schreiben', 'Verwalten', 'Kommentare', and 'Profil'. Below it, the 'Beiträge' tab is active. The form has a title field with the text 'Welche Suchmaschine ist die Beste?' and a 'Vorschau' button. Below the title is a text area with a rich text editor toolbar. The text area contains the placeholder 'Der Titel sagt es bereits..'. To the right of the text area is a 'Zusätzliche Kategorien' button. Below the text area is a 'Path' field. At the bottom of the form, there are buttons for 'Zwischenspeichern', 'Speichern', and 'Veröffentlichen'. Below these buttons is a 'Tags (kommasetrennte Liste)' field with a dropdown menu showing 'tit', 'titel', 'titel', 'titel', 'titel', 'titel'. To the right of this field is a 'Suggestive Tagging Funktion' button. Below the tags field is a 'Vorgeschlagene Tags' field with a dropdown menu showing 'ag', 'der', 'ei', 'gt', 'titel'. To the right of this field is an 'Automatisch generierte Vorschläge' button. On the right side of the form, there is a 'Kategorien' sidebar with a list of categories: 'Schule', 'Shopping', 'Sinnfrees', 'Software', 'Sport und Freizeit', 'Technik', and 'Telefon und Fax'. Below the categories sidebar are buttons for 'Diskussion', 'Passwort', 'Titelform', 'Status', and 'Zeitstempel'.

Abb. 6.4 Eingabemaske von Fragr

4.1.2 Yahoo! Clever

Die größte Auffälligkeit bei Yahoo! Clever ist, dass hier kein Tagging, sondern ein mehrstufiges System von Kategorien benutzt wird (vgl. Abb. 6.5). Der Benutzer kann

diese aus einer Auswahl von Oberkategorien und zwei Ebenen von Unterkategorien aussuchen. Bei einer Frage zu Autos wählt man dementsprechend die Oberkategorie “Autos, Transport & Verkehr”, danach die Unterkategorie “Autohersteller” und dann kann zwischen einzelnen Herstellern als letzte Unterkategorie gewählt werden. Ein kurzer Hilfetext neben dem Kategorienfeld soll dem Benutzer erklären, wozu die Kategorien gewählt werden können. Bei der Eingabe der Frage werden ähnliche Fragen angezeigt, die evtl. bereits eine Antwort haben.

Gesucht werden kann ebenfalls über die Kategorien. Es wird erst eine der 26 Oberkategorien gewählt, danach kann man die Suche durch zwei weitere Ebenen von Unterkategorien einschränken. Es ist auch möglich per Suchfeld zu suchen, wobei man in der erweiterten Suche zusätzlich Kategorien wählen kann in denen gesucht werden soll.



Abb. 6.5 Kategorien von Yahoo! Clever

4.1.3 wer-weiss-was

Dieser Dienst existiert bereits seit 1996 und benutzt kein Tagging. Anstelle dessen wird ein Forum benutzt, in welchem man seine Fragen stellen kann. Allerdings gibt es auch die Möglichkeit, andere Nutzer direkt anzuschreiben oder in einem Chat Fragen zu klären. Das Forum ist in verschiedene Bereiche gegliedert, z.B. Politik, Freizeit, Computer, die wiederum in Unterkategorien aufgegliedert werden. Um im Forum zu suchen, kann man entweder direkt in den einzelnen Bereichen oder über ein Suchfeld nach Artikeln im Forum suchen.

Bei der Registrierung, ohne die der Dienst nicht nutzbar ist, muss man verschiedene Themengebiete angeben, in denen man „Experte“ ist. Auf der Startseite

ist es dann möglich, in einem Eingabefeld nach Experten zu einem bestimmten Thema zu suchen und diese dann direkt mit einer Frage anzuschreiben. Die Chats sind ein Zusatzdienst von wer-weisswas und werden einige Tage vorher mit einem bestimmten Thema, z.B. „Neue Medien und Sicherheit an Schulen“ angekündigt und moderiert. Nutzer, die ihre Wissensgebiete passend zum Thema des Chats angegeben haben, werden zu diesem Chat als Experten eingeladen.

4.1.4 willwissen.net

In diesem Dienst werden, ähnlich wie bei Yahoo! Clever, keine Tags, sondern Kategorien verwendet. Die Kategorien, z.B. Sport und Sozialwissenschaften, werden dann noch in eine Reihe von Unterkategorien aufgegliedert. Bei der Suche kann man entweder direkt auf die Kategorien und Unterkategorien gehen oder über das Suchfeld suchen, wobei man auch hier eine Kategorie auswählen kann.

4.1.5 Quansr

Bei Quansr handelt es sich um einen weiteren Dienst, der sich noch in der Beta-Phase befindet und sich noch deutlich verändern kann. Quansr ist recht schlicht und unkompliziert gehalten. Da man sich nicht registrieren muss, gibt es auch keine Einschränkungen hinsichtlich der benutzbaren Funktionen, wie es bei anderen Diensten häufig der Fall ist, wenn man nicht registriert ist.

Auf der Hauptseite ist ein Fragefeld, unter dem sich ein Eingabefeld für Tags befindet. Es werden hier weder Fragen noch Tags vorgeschlagen. Geht man mit dem Mauszeiger auf das Fragezeichen, welches neben dem Eingabefeld für die Tags zu finden ist, wird kurz erklärt, wozu Tags benutzt werden. Es gibt bei Quansr kein Suchfeld sondern es kann nur über eine Tagcloud und über zwei andere, für das Tagging irrelevante, Möglichkeiten (letzte Fragen, letzte Antworten) gesucht werden, die es auch bei anderen Diensten gibt.

4.1.6 gutefrage.net

Auch dieser Dienst befindet sich noch in der Beta-Phase. Unter dem Link „F.A.Q.“ findet man einen Hilfetext, in dem erklärt wird, was ein Tag ist und wie es benutzt wird. Vor dem Speichern einer Frage erscheinen zudem noch Hinweise zum Tagging.

Bei der Eingabe der Frage werden andere Fragen angezeigt, in denen die gleichen Wörter vorkommen. Die zu diesen Fragen vergebenen Tags werden ebenfalls angezeigt. Suchen ist über ein Suchfeld und eine Tagcloud möglich. Anders als bei den meisten untersuchten Diensten wird hier bei der Suche über das Suchfeld keine kombinierte Suche durchgeführt. Dies bedeutet, dass bei der Eingabe von mehreren Suchwörtern nicht die Einträge angezeigt werden in denen alle der eingegebenen Wörter vorkommen, sondern es werden alle Einträge angezeigt, in denen mindestens eines der gesuchten Wörter vorkommt.

4.1.7 motorradfrage.net und sportlerfrage.net

Diese beiden Suchdienste bieten Fragen und Antworten zu einem speziellen Themengebiet an. Beide Dienste werden von gutefrage.net angeboten und sind dementsprechend damit fast identisch. Interessant ist allerdings die häufig bessere

Qualität der vergebenen Tags. Der Grund hierfür scheint das begrenzte Themengebiet zu sein, was bedeutet, dass weniger allgemeine Tags und mehr eindeutige Tags benutzt werden. So würde beispielsweise bei sportlerfrage.net das Tag „Sport“ zu allgemein sein.

4.2 Social Bookmarking-Dienste

Bei den in diesem Kapitel untersuchten Anbietern lassen sich URLs, also Bookmarks, speichern, verwalten und anderen Nutzern zur Verfügung stellen. Anders als Bookmark-Verwaltungen in Web-Browsern werden die Bookmarks nicht auf dem Computer, sondern auf dem Server des Dienstes gespeichert, womit sie von jedem Internetzugang aus verfügbar sind. Neben Diensten wie del.icio.us, Mister Wong, Furl, netselektor und oneview, die thematisch für alle Richtungen offen sind, stehen Dienste wie CiteULike, BibSonomy und Connotea zur Verfügung. Diese drei Anbieter sind in erster Linie für die Erfassung von wissenschaftlicher Literatur gedacht und bieten die Möglichkeit, nicht nur Online-Ressourcen, sondern auch Print-Quellen zu erfassen und zu verwalten. Da es bei den Social Bookmarking-Diensten sehr viele Ähnlichkeiten gibt, werden hier die Anbieter nicht einzeln vorgestellt, sondern es werden in diesem Kapitel die Gemeinsamkeiten und Unterschiede aufgezeigt. Um möglichst verschiedene Anbieter vorzustellen, wurden etablierte Anbieter (del.icio.us, Mister Wong) und weniger bekanntere Angebote (Furl, netselektor, oneview u.a.) gewählt. Die Auswahl wurde ergänzt mit Angeboten, die sich primär an ein wissenschaftliches Publikum richten (CiteULike, BibSonomy und Connotea).

Einen Überblick über alle untersuchten Anbieter gibt Abb 6.6, in der folgende Kriterien festgehalten wurden: Zunächst erfolgt ein Überblick über die verschlagwortete Ressource und den Tagging Support. Da es im Gegensatz zu den Frage-Antwort-Diensten für die User von einigen Social Bookmarking-Diensten möglich ist, Relationen zwischen Tags zu definieren, wurde dieses Kriterium mit in die Untersuchung aufgenommen. Ob bei den Suchergebnissen zusätzlich verwandte Tags – passend zum Suchbegriff – abgebildet werden, ist ebenfalls vermerkt.

Bei allen Diensten sind die Tagging-Rechte so definiert, dass jeder User seine Ressource selbst taggt und verwaltet („self-tagging“).

Anbieter	Ressource	Tagging Support	Definition von Relationen zwischen den Tags	Verwandte Tags in Suchergebnissen
Lycos iQ (Linkkatalog) iq.lycos.de	Bookmarks	viewable/suggestive	nein	ja
del.icio.us del.icio.us/	Bookmarks	viewable/suggestive	ja	ja
Mister Wong www.mister-wong.de	Bookmarks	viewable/suggestive	ja	ja
Furl www.furl.net	Bookmarks	suggestive	nein	nein
LinkARENA linkarena.com	Bookmarks	suggestive	nein	ja
netselektor www.netselektor.de	Bookmarks	suggestive	ja	ja
oneview www.oneview.de	Bookmarks	suggestive	nein	ja
BibSonomy www.bibsonomy.org	Bookmarks und Printartikel	viewable/suggestive	ja	ja
Connotea www.connotea.org	Bookmarks und Printartikel	suggestive	nein	ja
CiteULike www.citeulike.org	Bookmarks und Printartikel	suggestive	nein	nein

Abb. 6.6 Übersicht der untersuchten Social Bookmarking-Dienste

Darüber hinaus gibt es viele weitere Gemeinsamkeiten zwischen den Diensten, da einige Features scheinbar zu einem gewissen Standard gehören. So findet man jedes Mal die Unterstützung der User durch das „suggestive tagging“ bei der Tag-Eingabe. Die Quellen der vorgeschlagenen Tags unterscheiden sich. Es können Tags sein, die bereits von anderen Usern vergeben wurden oder nur von einem selbst. Oneview bietet darüber hinaus Vorschläge an, bevor überhaupt eine URL oder ein Tag eingegeben wurde. Diese Vorschläge sind dementsprechend sehr allgemein, wie z.B. Musik, Politik, Sport und Kultur, erinnern also eher an grobe Kategorien.

Netselektor greift dagegen auf die Begriffe anderer User zurück. Auffallend ist hier, dass während der Eingabe den vorgeschlagenen Tags kein alphabetischer Index zugrunde liegt und die Auswahl kleiner zu sein scheint als bei den anderen Anbietern. Während sonst bei der Eingabe der ersten zwei Buchstaben von „Auto“ auch tatsächlich dieses Wort vorgeschlagen wird, setzt netselektor vermutlich so etwas wie ein Stemming-Verfahren ein. Bei der Eingabe des Buchstabens „g“ werden beispielsweise „Kopfbedeckung“, „Übersetzung“ und „Zeugnis“ vorgeschlagen. Egal welchen Buchstaben man aus dem Wort „Kopfbedeckung“ eingibt, es wird jedes Mal angeboten. Dieser Umstand und dass die Auswahl der vorgeschlagenen Begriffe kleiner ist, lassen uns vermuten, dass es sich um Tags handelt, die besonders oft vergeben wurden. Ein weiterer und deutlicherer Hinweis darauf ist, dass sich viele der vorgeschlagenen Tags auch in der allgemeinen Tagcloud befinden. Teilweise bietet diese Art des „suggestive tagging“, die hier vorzufinden ist, aber auch interessante

Ergebnisse: Möchte man nun z.B. „Guide“ eingeben, werden einem auch etwa „Webguide“ oder „Festivalguide“ vorgeschlagen.

Im Eingabeformular bei del.icio.us können, wie bereits erwähnt, mehrere Möglichkeiten auftreten: Das System schlägt die populärsten Tags vor, die für die betreffende Ressource bisher von anderen Nutzern vergeben wurden („popular tags“). Die zweite Variante sind die „recommended tags“, also Vorschläge des Systems, die sich danach richten, welche Tags man selbst – oder andere Nutzer – bisher genutzt haben.

Diese Variante, bei der dem User im Eingabeformular die Tags angezeigt werden, die bereits für die URL vergeben wurden, das sogenannte „viewable tagging“, wird neben del.icio.us und Mister Wong auch von BibSonomy angeboten.

Sieben der neun Social Bookmark-Dienste geben bei ihren Suchergebnissen auch verwandte Begriffe an, was z.B. auch eine Lösung sein kann, um bei einer Recherche Synonyme oder Singular- bzw. Pluralformen mit einzubeziehen. Die Qualität der verwandten Tags variiert bei den Systemen unserer Erfahrung nach etwas, was aber vermutlich auch mit der Anzahl der Nutzer und der Menge der Daten bzw. Tags zu tun hat.

Beispielsweise gibt Mister Wong bei seinen Suchergebnissen auch verwandte bzw. synonyme Begriffe an, mit denen sich die Suche fortführen lässt. Wurde der Suchbegriff „auto“ gewählt, werden – zusätzlich zu den Treffern für „auto“ – am rechten Bildrand die Tags „autos“, „kfz“, „pkw“ oder „gebrauchtwagen“ angezeigt (vgl. Abb. 6.7).

The screenshot shows the Mister Wong website interface. At the top, there is a search bar with the text 'auto' and a 'suchen' button. To the right of the search bar, there is a section titled 'News im Blog' with the text 'Die Januar Charts' and 'Das neue Logo von Mister Wong'. Below the search bar, there is a navigation menu with links: Übersicht, Charts, Spotlight, Registrieren, Einloggen, Gruppen, Stuff, Hilfe, Blog. To the right of the navigation menu, there is a counter showing '3.504.096'. Below the navigation menu, there is a section titled 'Google Anzeigen' with two ads. The first ad is for 'Auto' with the text 'Infos zu aktuellen Modellen & mehr Ford - Feel the difference' and 'www.ford.de'. The second ad is for 'Günstige Kfz-Reparatur' with the text 'Europas führende Gruppe freier Werkstätten. Alle Marken, alle Arbeiten' and 'www.ad-autodienst.de'. Below the ads, there is a section titled 'auto' with a 'sehr beliebt' button and a 'brandneu' button. To the right of the 'auto' section, there is a list of related tags: 'verwandt mit', '+ kfz', '+ tuning', '+ car', '+ autos', '+ gebrauchtwagen', '+ motorrad', '+ blog', '+ neuwagen', '+ forum', '+ pkw', and '...'. Below the related tags, there is a section titled 'Spritverbrauch berechnen und Autokosten verwalten' with a link to 'vergleich auto sprit spritmonitor spritverbrauch benzin verbrauch benzinverbrauch tanken tankkosten benzinpreis kfz' and a button 'Hinzugefügt vor >30 Tagen von cf, 62 Benutzer speichern'. At the bottom right, there is an advertisement for 'ANIMAX Serien PREMIERE FUSSBALL'.

Abb. 6.7 Anzeige verwandter Begriffe

Bei netselektor werden neben den verwandten Tags auch Prozentzahlen für ihre Relevanz für die Suche angegeben. Sucht man nach dem Tag „fußball“, wird für Tags wie „shaolin“, „fu fighters“ oder „universität“ allerdings eine Relevanz zwischen 71 und 78 Prozent für die aktuelle Anfrage angegeben, obwohl sie unpassend sind.

Um die Suche gezielter durchführen zu können, besteht bei den meisten Diensten die Möglichkeit, mit mehr als einem Tag zu suchen und so die Treffermengen zu reduzieren. Entweder gibt man dafür die Tags einfach nacheinander in die Suchmaske ein oder man hat die Option, wie bei oneview oder Mister Wong, Tags mit einem Pluszeichen zu kombinieren. Bei CiteULike benutzt man Booleschen Operatoren „AND, OR, NOT“ und kann Suchbegriffe auch trunkieren. Auch bei del.icio.us kann man diese Operatoren einsetzen. Als einziger Dienst bietet netselektor keine kombinierte Suche mit mehreren Tags an.

Bei allen neun Diensten können die User ihre bereits vergebenen Tags editieren, also umbenennen oder löschen. Bei Mister Wong z.B. lassen sich Tags in einem Schritt für alle Ressourcen ändern (vgl. Abb. 6.8). Dies ist sinnvoll, wenn sich die „Tagging-Skills“, also die Fähigkeit zu Taggen, mit der Zeit und wachsender Erfahrung verbessern.

Tags editieren

Tags umbenennen

Altes Tag:

Neues Tag: (maximal zwei, durch Leerzeichen getrennt)

Tags löschen

Tag:

Abb. 6.8 Editierung von Tags bei Mister Wong

Bei allen Diensten sind nützliche Hilfsfunktionen mehr oder weniger gut sichtbar eingebunden, so dass auch unerfahrene User schnell die Funktionsweise und den Sinn des Taggings begreifen. BibSonomy und Connotea bieten zudem Video- bzw. Flash-Tutorials an.

Die beliebtesten Tags werden bei acht der neun Dienste in Tagclouds angezeigt. Die Tagclouds unterscheiden sich insofern, dass sie für einen einzelnen User oder die komplette Gemeinschaft gelten. Oneview zeigt zudem in der „Trend-Wolke“ die Häufigkeit der Tags in Klammern an. Bei Furl werden zwar „hot topics“, also beliebte Tags angezeigt, aber nicht in der gewohnten Visualisierung, bei der die beliebtesten Tags am größten dargestellt werden.

Einige Features sollten hier noch hervorgehoben werden, da sie nur von einzelnen Diensten angeboten werden. Die User von BibSonomy, netselektor, Mister Wong und del.icio.us haben die Möglichkeit, ihre Tags in Ordnern zu verwalten bzw. Relationen festzulegen. Bei Mister Wong und del.icio.us hat der User die Möglichkeit sogenannte „Tagbundles“ (Tag-Bündel) anzulegen. Dieses System ähnelt vom Prinzip

her einem Ordnersystem, wie man es von Microsoft Windows kennt, um dort seine Dateien zu ordnen. So werden im Grunde Ober- und Unterbegriffe, also Beziehungen zwischen den Begriffen, definiert. In einem Ordner „Software“ kann der User Links bzw. Tags speichern, wie „Excel“, „Dreamweaver“ usw. Zum einen lassen sich Tags so besser und übersichtlicher verwalten, zum anderen erkennt das System automatisch Zusammenhänge zwischen den indexierten Wörtern. Inwieweit diese Informationen genutzt werden, um z. B. zu einem Suchbegriff passende verwandte Begriffe („related tags“) anzuzeigen, konnte nicht wirklich nachvollzogen werden. Nach verschiedenen Versuchen kam man zum Ergebnis, dass sich die „related tags“ vermutlich eher aus einer Tag-Gruppe zusammensetzten, die ein User zusammen mit dem betreffenden Tag besonders häufig für eine Quelle vergeben hat. Die Verwaltung mithilfe von Ordnern bietet netselektor ebenfalls an.

Bei BibSonomy kann man die Beziehungen über Pfeile zwischen den Tags herstellen und so Ober- und Unterbegriffe definieren. Dazu verbindet man in die Eingabezeile die Tags mit Pfeilen und definiert sie somit als Ober- und Unterbegriffe (vgl. Abb. 6.9).

BibSonomy :: edit bookmark all <fulltext search

A blue social bookmark and publication sharing system.

tags · groups · relations · popular
myBibSonomy · post bookmark · post bibtex · myRelations

logged in as example · hel
0 picked in basket · edit tags · frier

Feel free to edit your bookmark

url*

title*

description, comment

tags*

recommendation: google internet search_engine dblp googles test

suggested

viewable for public save

tags
list | cloud
bibtex example
openurl p2p publi
alph | freq

With this arrow, you can define relations between your tags.
Here: google is a search_engine.

Abb. 6.9 BibSonomy-Tutorial: Herstellen von Beziehungen zwischen Tags

Eine Suche nach diesen von den Usern definierten Beziehungen ist ebenfalls möglich. Zudem gibt es eine Übersicht über die häufigsten Relationen, die vom Prinzip her an eine Tagcloud angelehnt ist (vgl. Abb. 6.10).

BibSonomy :: relations ▼ ::

A blue social bookmark and publication sharing system.

tags • relations • groups • popular
myBibSonomy • post bookmark • post bibtex

logged in as geilmann • help • blog • about
0 picked in basket • edit tags • settings • logout

Here you can see the most popular relations of our users. You can access your own relations via myRelations and edit them on the edit tags page.

- AI — agent agents ai MBR planning QR scheduling
- bibliothek — bibliothekssysteme nationalbibliothek usb-koeln web2.0
- Bookmarks_Toolbar_Folder** — Bibsonomy Binproject Blue_Dot Bookmarks_Portals CafeH Ent_Lib FaceBook forex-directories Furl Geotaching Getting_Started GIS Home Imported Links linux LIVE Live_Bookmarks MPR MST-Dortmund NetCologne Netscape_bookmarks News Opera PC PIT promotion Publishers_and_Commissions reg-directories REO rözne SAD simply Sprachen studia Suchmaschinen survey Temp temp Thongkie tm Uni Yahoo! Zeitungen
- Business — Competitive_Intel Customer_Service Domains Email_Mgmt Keywords Legal Networking News_I_Can_Use Niche_R&D Office_Supplies Product_Creation Real_Answers Security SEO Word_of_Mouth
- clustering — biclustering cbc cluster clusterm co-clustering coclustering folksonomy hac kmeans multilevel social_network socialnets
- computer — layouts linux Mac source_code workstation
- Documentation — Dépôt_PDF Dictionnaires Gazetteer moteur_de_recherche Portail repository RSS Thesaurus veille
- eclipse — eclipse_plugins eclipse-help eclipsepedia eclipsewiki emf gef gmf yxos
- entertainment — entertainment.books entertainment.comics entertainment.demos entertainment.humor entertainment.legacy entertainment.radio
- Film — brabetyfilm DVD_Fernsehen_-_Technik Film_Kritiken Filme Filmportale_-_DVD_Reviews Filmrecension filmrecension_dokumentär filmtips-udda intressantafilmer Publisher Shop TV_Serien
- folksonomy — bibsonomy communities flickr_-_photosharing folkrank knowledgeorganization mash-up_with_googlemap media_web_editing participation resources search_feed slide_sharing social tagcloud tagging twitter_-_nanopublishing video_sharing Web2.0 wiki

Abb. 6.10 Die beliebtesten Relationen

Das Feature „tag description“ bietet die Option, die eigenen Tags zu beschreiben. Del.icio.us erklärt diese Funktion mit dem Beispiel-Tag „Zlorn“, das ein Phantasiewort des Users osunick zur Beschreibung von besonders interessanten Internetseiten ist [5]. Der Nutzer erstellt also eine Definition für dieses Tag, welche für alle User sichtbar wird. Benutzen einzelne User besonders eigenwillige Tags, die für andere User nicht nachvollziehbar sind, so ist dies eine Lösung, damit alle anderen User den Sinn nachvollziehen können. Connotea bietet ebenfalls eine „Add Tag Note“-Funktion an, bei der Erklärungen an die Tags angefügt werden können.

Erwähnenswert ist noch der Versuch, den Connotea mit dem Plug-In EntityDescriber gestartet hat. Es handelt sich hier um eine Software, mit der sich ein kontrolliertes Vokabular für das Tagging nutzen lässt. Ähnlich wie bei dem „suggestive tagging“ hat der Nutzer hier eine Funktion, die bei der Eingabe die Worte automatisch vervollständigt [6]. Als verfügbare kontrollierte Vokabulare sind „MeSH“, „Atom ontology“ und „Person ontology“ aufgelistet.

4.3 Andere Web-Anwendungen

In diesem Kapitel finden sich eine Reihe unterschiedlicher Dienste, die neben den Social Bookmarking- und den Frage-Antwort-Diensten untersucht wurden. Abb. 6.10 gibt eine Zusammenfassung der Dienste, die im Folgenden vorgestellt werden. Mit der Auswahl der Dienste soll ein grober Überblick über andere Web-Anwendungen gegeben werden, welche ebenfalls Tagging benutzen. Dabei wurde darauf geachtet, dass möglichst unterschiedliche Ressourcen verschlagwortet werden oder ein Dienst Innovatives zu bieten hat, wie die Einbindung vom Tagging in einen Bibliothekskatalog.

Anbieter	Dienst	Ressource	Tagging Support	Tagging Rights
Flickr www.flickr.com	Fotoportal	Fotos	blind	permission-based
Last.fm www.lastfm.de	Musikportal	Musik (Künstler, Songs etc.)	viewable	free for all
YouTube www.youtube.com	Videportal	Videos	blind	self-tagging
MyVideo www.myvideo.de	Videportal	Videos	blind	self-tagging
Amazon.com www.amazon.com	Versandhandel-Seite	diverse Produkte (u.a. Bücher, Musik)	suggestive	free for all
43things www.43things.com	Social Networking-Portal	Tätigkeiten, die man sich für die Zukunft vornimmt	viewable	self-tagging
Kölner UniversitätsGesamtkatalog kug.ub.uni-koeln.de	Bibliothekskatalog	Medien aus dem Bibliothekskatalog	blind	free for all

Abb. 6.10 Übersicht der sonstigen Web-Anwendungen

4.3.1 Amazon.com

Amazon.com ist einer der bekanntesten Online-Versandhändler. Angefangen mit Büchern können inzwischen sogar Kleidungsstücke bei Amazon erstanden werden. Aufgrund der Struktur und Art dieser Webpage ist das Erstellen eigener Beiträge für die Nutzer nicht möglich. Verschlagwortet werden also, im Gegensatz zu den meisten anderen Seiten, die untersucht wurden, nicht eigene Beiträge, sondern die von Amazon.com angebotenen Produkte.

Das Tagging ist auch nicht Pflicht, wie bei den meisten anderen Diensten, sondern freiwillig. Dies und die versteckte Platzierung der Tagging-Funktion zwischen den „Editorial Reviews“ und den „Customer Reviews“ lassen den Schluss zu, dass das Tagging bei Amazon.com nur ein Feature ist, und nicht sonderlich wichtig. Trotzdem gibt es einen sehr ausführlichen Hilfetext auf der Hilfeseite.

Beim Eingeben der Tags in das Eingabefeld gibt es eine Funktion, die Tags nach deren Nutzungshäufigkeit vorschlägt (vgl. Abb. 6.11). Gibt man z.B. „Car“ ein, erscheint ein Fenster, in dem man sehen kann, dass Cartoon mit 1362 Mal das am häufigsten benutzte Schlagwort mit den eingegebenen Elementen ist und Car 609 Mal verwendet wurde. Für die Suche nach Tags kann man nicht das große Suchfeld am Kopf der Seite benutzen, da in diesem Feld nur nach Titel, Author usw. gesucht wird. Stattdessen gibt es in dem Abschnitt auf der Produktseite, in dem man auch die Tags eingeben kann, ein weiteres, kleines Suchfeld in welchem man nach Tags suchen kann. Desweiteren gibt es eine Tagcloud, in der man ebenfalls nach Tags suchen kann.

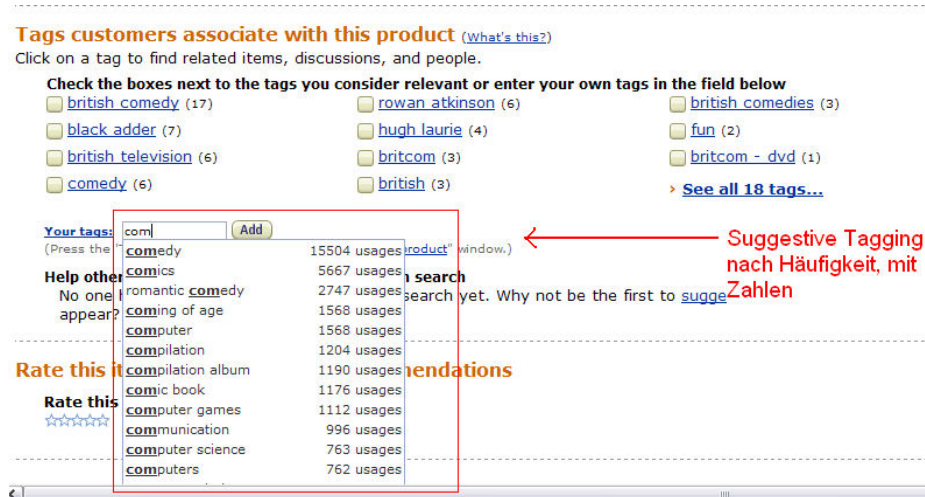


Abb. 6.11 „Suggestive tagging“-Funktion bei Amazon.com

Klickt man auf ein Tag, werden alle Produkte angezeigt, die mit diesem Tag versehen wurden. Diese Produkte werden dann nach Häufigkeit der Tags geordnet und man hat die Möglichkeit, über einen Button „+“ das Produkt ebenfalls mit diesem Tag zu versehen, sowie mit einem Button „-“ wieder zu löschen bzw. der angeschlossenen Amazon-Community zur Löschung vorzuschlagen (vgl. Abb. 6.12).

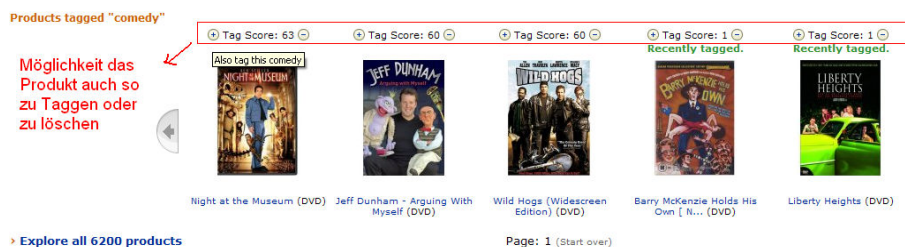


Abb. 6.13 Tagsuche bei Amazon.com

4.3.2 43 Things

43 Things ist eine Social Networking-Seite, also eine Webseite auf der sich die Nutzer miteinander verknüpfen können. Hier kann man Dinge, die man sich vornimmt, oder Ziele, die man erreichen möchte, auf der Website eintragen und nach den Zielen anderer bzw. nach Dingen, die man machen möchte, suchen. So kann ein User, der mit dem Rauchen aufhören möchte, Gleichgesinnte finden. Ein weiteres Beispiel ist „Go Skydiving“: Alle Leute, die gerne mal Fallschirmspringen möchten, oder dies bereits als Hobby haben, können sich ebenfalls für dieses Ziel eintragen. So kann man sich mit anderen Nutzern dieser Seite verknüpfen.

Es werden die Dinge, die man gerne machen würde, mit Tags versehen, allerdings sind Tags hier nicht unbedingt notwendig und werden dementsprechend auch nicht

von allen Nutzern verwendet. Die Tags werden für die Suche über das Suchfeld benutzt, um die Themen, die meist aus mehreren Wörtern bestehen, zu finden.

Wenn man ein Thema bzw. sein Ziel fürs Leben, auswählt, kann man sehen, welche Tags am häufigsten für dieses Thema vergeben wurden. Beispielsweise wird beim Thema „Travel the World“ angezeigt, dass u.a. „adventure, travel, fun, world“ beliebte Tags sind. Unter dieser Anzeige ist eine Schaltfläche mit einem Eingabefeld zu finden, in welches man selbst ein Tag für dieses Thema eingeben kann.

4.3.3 Kölner UniversitätsGesamtkatalog

Bei dem Kölner UniversitätsGesamtkatalog handelt es sich um den Onlinekatalog der Universität Köln. Hier können Medien gesucht, bestellt und verlängert werden, die in den Universitätsbibliotheken ausgeliehen werden können. Ähnlich wie bei Amazon.com werden hier keine von den Nutzern erstellten Beiträge verschlagwortet, sondern Medien des UniversitätsGesamtkatalogs. Auf den Hilfeseiten des Gesamtkatalogs befinden sich mehrere Tutorials, die als PDF-Datei verfügbar sind, in denen auch das Tagging erklärt wird.

Gibt man Tags im Eingabefeld ein, werden die Wörter automatisch klein geschrieben und Umlaute aufgelöst. Es gibt die Möglichkeit, die Tags entweder „privat“ nur für einen selbst anzeigen zu lassen oder für alle Nutzer sichtbar zu machen. Man kann über den Link „Tags“ eine Tagsuche starten. Die Tags dort werden wie in einer Tagcloud dargestellt, klickt man auf ein Tag, bekommt man alle so verschlagworteten Einträge angezeigt.

4.3.4 MyVideo

Bei MyVideo handelt es sich um ein Videoportal. Nutzer können Videos auf die Seite stellen und sich Videos anderer ansehen. Außerdem kann man Beiträge kommentieren und bewerten, sowie Interessengruppen beitreten. Es werden die Videobeiträge von den Nutzern, die diese Beiträge erstellen, mit Tags gekennzeichnet.

Auf der Seite mit der Eingabemaske für die Videos befindet sich unter dem Feld für die Tags eine kurze Erklärung, wie man dort die Tags eingibt. Unter dem Link „Hilfe“ findet sich unter dem Unterpunkt „allgemeine Fragen“ ein etwas ausführlicherer Text, der die Funktion der Tags erklärt. Der Nutzer kann ein oder mehrere Schlagwörter in einem Eingabefeld eingeben. Die Schlagwörter werden mit einem Komma getrennt. Es werden keinerlei Vorschläge für Tags gemacht. Ähnlich wie es bei Fragr der Fall ist, kann man sich zusätzlich zu seinen Schlagwörtern zwischen einigen Kategorien für seinen Beitrag entscheiden. Anders als bei Fragr ist die Auswahl einer Kategorie hier allerdings Pflicht und es kann auch nur eine Kategorie gewählt werden. Es gibt 24 Kategorien, z.B. „Auto & Verkehr“.

Für die Suche gibt es mehrere Wege: Zum einen kann man über ein Suchfeld suchen, zum anderen ist eine Suche über Klick auf die verschiedenen Kategorien möglich. Desweiteren kann man über eine Tagcloud mit den meist verwendeten Tags suchen. Die Suche per Suchfeld sucht sämtliche beim Erstellen eines Beitrages angegeben Informationen ab, d.h. es wird im Titel, in der Beschreibung, in den Schlagwörtern und in den Kategorien gesucht. Werden mehrere Wörter eingegeben, zeigt das Suchergebnis nur Beiträge an, bei denen alle eingegebenen Wörter gefunden wurden. Klickt man auf eine Kategorie, werden erst einmal alle Beiträge, die in dieser

Kategorie vergeben wurden, bei 20 Beiträgen pro Seite, angezeigt. Dies ist natürlich bei den vielen Tausend Videos, die pro Kategorie existieren, etwas unübersichtlich. Klickt man auf ein Tag in der Tagcloud, z.B. „Fußball“, bekommt man alle so verschlagworteten Beiträge angezeigt. Zusätzlich dazu kann der User noch unter den Kategorien, z.B. „Hobbys“, auswählen, dann werden nur die Beiträge angezeigt, die „Fußball“ als Tag haben und in der Kategorie „Hobbys“ eingetragen wurden.

4.3.5 YouTube

YouTube ist das wahrscheinlich bekannteste Videoportal. Wie bei MyVideo können Nutzer selber Videos auf die Seite stellen und sich Videos anderer ansehen. Auch hier kann man Beiträge kommentieren und bewerten, sowie Interessengruppen beitreten. Der Inhalt, der verschlagwortet wird, sind auch hier die Videos; die Tags werden vom Ersteller eines Beitrages vergeben.

Als erklärende Hilfe gibt es unter „Hilfe“ einen Text, der die Funktion der Tags erklärt, zudem gibt es in der Eingabemaske über dem Feld für die Tags eine kurze Erklärung. Die Tags werden in ein Eingabefeld eingegeben und durch einen Leerschritt getrennt. Wie es auch bei MyVideo der Fall ist, werden hier ebenfalls keinerlei Vorschläge für Tags gemacht, aber auch hier kann man zusätzlich zu den Tags eine Kategorie wählen. Die Wahl einer Kategorie ist ebenfalls Pflicht. Es gibt 14 Kategorien, z.B. „Autos & Vehicles“.

Suchen sind über ein Suchfeld und per Klick auf die Kategorien möglich. Eine Tagcloud gibt es nicht. Die Suche über das Suchfeld ist vergleichbar mit der Suchfeld-Suche bei MyVideo. Gibt man ein Suchwort ein, werden alle beim Erstellen des Beitrages eingegebenen Informationen abgesucht, also der Titel, die Beschreibung, die Tags und die Kategorien. Werden mehrere Wörter eingegeben, zeigt das Suchergebnis nur Beiträge an, bei denen alle eingegebenen Wörter gefunden wurden. Wie bei vielen anderen Diensten gibt es auch hier noch andere, nicht für das Tagging relevante Suchmöglichkeiten. Bei YouTube kann man diese Suchmöglichkeiten mit der Kategoriensuche kombinieren. Es gibt also noch eine Suche nach zeitlichen Aspekten (diese Woche, diesen Monat) und eine Suche nach anderen Optionen (neueste, meist gesehene, top bewertete). Diese Suchen werden mit den Kategorien kombiniert (vgl. Abb. 6.14). Dies bedeutet, dass man, wenn man eine Kategorie anklickt, auch immer eine der zeitlichen und eine der anderen Optionen auswählt. Die Grundeinstellung hier ist „Meist gesehene“, „Heute“ und für die Kategorien „Alle“.

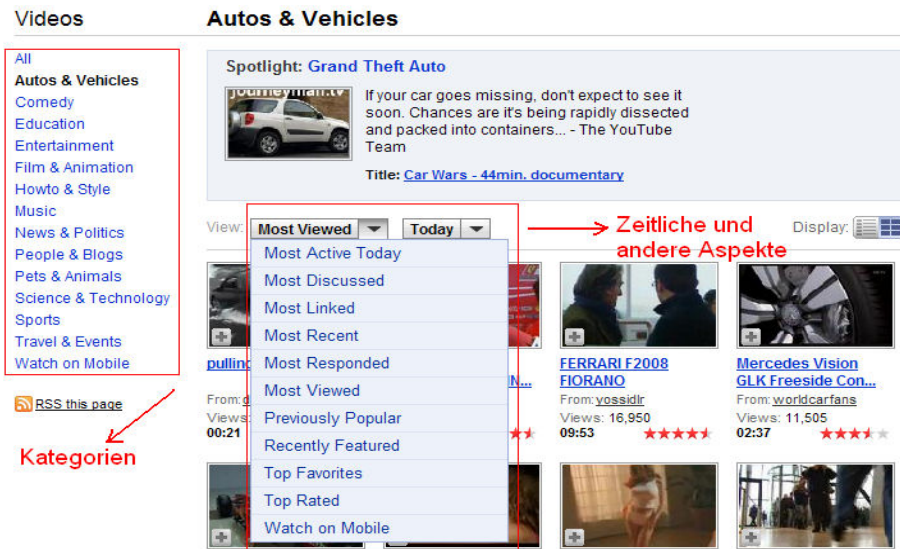


Abb. 6.14 Kombination von Kategoriensuche und anderen Aspekten bei YouTube

4.3.6 Flickr

Bei Flickr können die User ihre selbst gemachten Fotos hochladen und online verwalten. Die Tags dienen, um ihre Fotos zu beschreiben.

Wie auch z. B. alle Social Bookmarking-Dienste bietet Flickr die Möglichkeit, Tags zu bearbeiten, zu löschen oder per Tagcloud zu suchen. Im Unterschied zu den anderen von uns untersuchten Diensten werden die Tagging-Rechte für Fotos hier allerdings von dem User festgelegt, der die Fotos hochlädt („permission-based“). Er entscheidet, ob auch User aus seiner Freundesliste – oder sogar alle Flickr-Nutzer – die Fotos taggen dürfen.

Eine besondere Lösung für die Problematik, die beim freien Tagging durch Homonyme und Synonyme entsteht, bietet das in Flickr integrierte Feature „Cluster“, bei dem Tags automatisch gefiltert werden. Gibt man bei der Cluster-Suche beispielsweise das Tag „Apple“ ein, erhält man die Suchergebnisse in drei Abschnitte untergliedert (vgl. Abb. 6.15). Flickr erkennt drei unterschiedliche Bedeutungen des Wortes „Apple“ und unterteilt die Treffer in drei Cluster:

- die Frucht (englisch: „apple“ für „Apfel“)
- den Macintosh-Computer vom Hersteller Apple
- die als „Big Apple“ bezeichnete Metropole New York

Die Ergebnisse werden also in diese drei Gruppen unterteilt. Innerhalb dieser einzelnen Gruppen werden wiederum Synonyme, verwandte oder weitere Begriffe angezeigt, die teilweise auch die Pluralform des Suchbegriffs oder die Übersetzung in eine andere Sprache beinhalten.

Die Frucht wird mit den Begriffen *“fruit, red, food, green, orange, apples“* in Verbindung gebracht, für den Mac-Rechner gibt Flickr *“mac, macintosh, ipod“* usw.

aus. Die gefilterten Synonyme für “Big Apple“ lauten “*nyc, newyork, newyorkcity, manhattan*”.



Abb. 6.15 Suche mit Cluster-Funktion bei Flickr

4.3.7 Last.fm

Last.fm ist ein personalisiertes Internetradio, das seinen Nutzern auf Basis ihrer Hörgewohnheiten neue Musik, Menschen mit ähnlichem Musikgeschmack und Konzerte in ihrer Umgebung empfiehlt.

Die User taggen Künstler, also Musiker, Sänger und Bands sowie Musikstücke, die sie gerne hören. Der thematische Umfang ist im Gegensatz zu Lycos iQ also eher limitiert, da sich viele Tags einfach auf Genres beschränken. Außerdem gibt es Tags wie „seen live“, wobei das Tag auf einen Konzertbesuch des Users bei dem getaggtten Künstler hinweisen soll.

Last.fm bietet dem User einen gut verständlichen Hilfe-Text, der erklärt, was Tags sind, wozu sie dienen und wie man am besten taggt. Der User kann sehen, wie andere User die Musik verschlagwortet haben. Man kann sämtliche bereits bestehende Tagclouds zu einem Künstler einsehen und sie ergänzen.

Musik suchen

Künstler Alben Titel Tags Labels

rock

Suchen

Tags

27.489 Tags passende Treffer für "rock"

1 2 3 4 5 6 7 8 9 10 11 Nächste >

Tag	Beliebtheit	Tag	Beliebtheit
rock		prog rock	
classic rock		country rock	
hard rock		college rock	
progressive rock		experimental rock	
alternative rock		rock and roll	
indie rock		modern rock	
punk rock		celtic rock	
soft rock		instrumental rock	
garage rock		death rock	
stoner rock		rock opera	
psychedelic rock		dance rock	
southern rock		geek rock	
space rock		jazz rock	
post rock		math rock	
christian rock		electronic rock	
industrial rock		rock n roll	
pop rock		british rock	
blues rock		brit rock	
acoustic rock		symphonic rock	
gothic rock		funk rock	
folk rock		cello rock	
art rock		desert rock	

Abb. 6.16 Suchergebnisse in Last.fm

Werden bei der Suche Tags als Treffer angezeigt, findet sich auch eine Angabe über ihre Beliebtheit (also Häufigkeit) bei der Nutzung durch andere User. Zudem werden verwandte Begriffe angezeigt bzw. bei einem sehr allgemeinen Suchbegriff wie „Rock“ auch Kombinationen mit anderen Wörtern (vgl. Abb. 7.12).

5. VORSCHLÄGE FÜR LYCOS IQ

Lycos iQ bietet gegen die Probleme des freien Taggings (vgl. Kap. 2) bereits gängige Lösungen an wie das Anbieten von Vorschlägen durch das „suggestive tagging“ oder das Anzeigen von verwandten Begriffen. Nach der Analyse des Taggings anderer Dienste sollen hier nun einige Möglichkeiten zur weiteren Verbesserung des Taggings in Lycos iQ aufgezeigt werden, die auf anderen Seiten umgesetzt werden oder zumindest neue Ideen andeuten könnten. Diese Vorschläge sind als Anregungen gedacht und sollen zeigen, welche Mittel zur Verbesserung unserer Meinung nach denkbar sind.

5.1 Erweiterung der Tagging-Rechte

Bei Lycos iQ wurde für die Nutzer das Tagging-Recht so definiert, dass nur die Person, die eine Frage stellt, Tags erstellen kann („self-tagging“).

Es wird vorgeschlagen diese Variante abzuändern und stattdessen die Rechte zu erweitern. Es sollte auf „permission-based“ umgestellt werden, was bedeuten würde, dass der User beim Einstellen der Frage entscheidet, ob auch Leute, die antworten, die Frage bzw. ihre selbst gegebene Antwort taggen dürfen. Eine weitere Möglichkeit wäre, die Erlaubnis für Freunde oder gar alle User zu erteilen oder User für ihre Fachgebiete, also bestimmte Themen, freizuschalten.

Der so entstehende Vorteil wäre, dass die Antwort bzw. die Frage mit zusätzlichen Tags ergänzt wird, so dass zusätzliche Aspekte (die bei der Fragestellung nicht klar waren bzw. die erst in der Antwort enthalten sind) hier als relevante Suchbegriffe bzw. Tags im Nachhinein mit eingeführt werden können. Auf diese Weise können die User also der Tagging-Auswahl zu einem größeren Umfang verhelfen und damit die Auffindbarkeit verbessern.

Eine Frage zum Umformen von Video-Dateien des Users Lukas93 ist beispielsweise getaggt mit „pc“, „computer“ und „audio“. Die Antworten enthalten Informationen zu Funktionen von den Softwareprogrammen Nero und Audacity. Die Person, die antwortet, könnte in diesem Fall theoretisch also zusätzliche Tags wie „Nero“ und „Audacity“ angeben, da die sich nun erst aus seiner Antwort ergeben. Zudem hätte der User die Frage, wenn er sie gestellt hätte, mit „Konvertierung“ verschlagwortet, also fügt er dieses Tag zusätzlich mit an.

Seine Frage zu Steuerbefehlen auf seiner Tastatur hat User Ventavox mit „Strg“ getaggt. Die Frage bzw. die Antwort könnte durch den zweiten User mit „Steuerung“ und „Tastatur“ oder „Tastaturbefehle“ ergänzt werden. Die Varianten werden vielseitiger und die Chance der Auffindbarkeit steigt. Zudem können durch weitere Nutzer Begriffe eingefügt werden, die synonym verwendet werden.

Eine Möglichkeit, Rechtschreibfehler zu korrigieren, wäre hier zusätzlich gegeben. Gibt ein User nur das einzelne Tag „desktop“ ein, wird die Frage kaum bei einer Recherche gefunden werden. Ergänzt der antwortende User die Frage bzw. seine Antwort mit „desktop“, wird die Frage leichter auffindbar. Das Tagging durch weitere Nutzer hätte also eine Kontrollfunktion.

Der User, der eine Frage stellt, erhielte so auch ein Feedback zu seinen vergebenen Tags und könnte unter Umständen diese „Verbesserungsvorschläge“ für zukünftige Fragen übernehmen. Hier könnte also ein Lerneffekt generiert werden.

5.2 Verbesserte Editiermöglichkeit der Tags

Sinnvoll ist die Möglichkeit, seine selbst vergebenen Tags bearbeiten zu können, da sich im Laufe der Zeit die „Tagging-Skills“ verbessern. Diese Bearbeitungsoption ist zwar auch im Bookmark-Dienst von Lycos iQ integriert. Allerdings bieten Mister Wong und del.icio.us hierfür eine bedienerfreundlichere Variante an, zumindest wenn eine große Anzahl von Bookmarks gespeichert worden ist. Verwaltet ein User beispielsweise 200 gespeicherte URLs, die mit dem Tag „www“ versehen sind, und möchte er nun bei jedem Bookmark dieses Tag durch „Internet“ ersetzen, müsste er jedes Bookmark einzeln bearbeiten.

Unter „Tags editieren“ hat man bei Mister Wong einen Index mit allen von einem selbst vergebenen Tags. Hier wählt man z.B. das Tag „bloggen“ aus und kann es für alle Ressourcen in einem Schritt in „blog“ ändern. Gerade bei großen Mengen an Bookmarks ist dies eine große Erleichterung.

Ein weiteres nützliches Feature ist das Anhängen einer Definition an ein vergebenes Tag, wie es sich bei del.icio.us („tag description“) und Connotea („Add Tag Note“) machen lässt. Der Nutzer kann hier eine Definition für seine verwendeten Tags anfügen, die daraufhin für alle User sichtbar wird. Benutzt jemand also besonders eigenwillige Tags, die für andere User nicht nachvollziehbar sind, so ist dies eine Lösung, damit alle anderen User den Sinn nachempfinden können.

5.3 Cluster-Funktion

Die Cluster-Funktion bei Flickr ist sehr beeindruckend, da ihre Ergebnisse sehr nützlich sind und Homonyme, Synonyme, Sprachen und verschiedene Schreibweisen oder Numeri effektiv für die Suche der Nutzer gefiltert und gebündelt werden. Inwieweit es möglich ist, ein ähnlich funktionierendes System für Lycos iQ bzw. allgemein für andere Dienste umzusetzen, kann allerdings schwer beurteilt werden.

5.4 Motivation

Die Motivation der User zum Taggen ist schwer einzuschätzen. Es ist zu vermuten, dass ein gewisser Teil der User in erster Line die Frage beantwortet haben möchte und einfach deshalb Tags eingibt, da dies notwendig ist, um die Frage zu speichern. Über die spätere Wiederauffindbarkeit macht er sich unter Umständen wenig Gedanken. Würde ein Anreiz geschaffen, damit die Fragen möglichst gut recherchierbar sind, könnte sich das Tagging bei dieser Gruppe von Usern unter Umständen verbessern. Ein Anreiz könnte beispielsweise dadurch geschaffen werden, indem der User sieht, wie oft (auch über einen längeren Zeitraum) seine Frage aufgerufen wird. Für hohe Aufrufzahlen könnte man Punkte vergeben. Auf der anderen Seite spielt sicherlich nicht die Recherchierbarkeit, sondern die Interessantheit der Frage eine größere Rolle.

5.5 Verwaltung der Tags

Die „Tag-Bündel“ (vgl. Kap. 4.2) bei Mister Wong oder del.icio.us sind ein gutes Beispiel für die optimierte Verwaltung der Tags, die auch für die User des Social Bookmark-Dienstes von Lycos iQ interessant sein könnten. Der User schafft mit dem Einsatz von Ordnern bessere Strukturen und mehr Übersicht über seine Tags. Zudem lassen sich diese vom User definierten Beziehungen für Suchergebnisse nutzen. Etwa durch die Anzeige verwandter Begriffe.

5.6 „Suggestive tagging“-Funktion

Ein Vorschlag, der vermutlich relativ einfach umsetzbar wäre, ist, die „suggestive tagging“-Funktion noch ein wenig zu verbessern, indem man (ähnlich wie bei Amazon.com) zusätzlich die Häufigkeit der Tags in Form von Zahlen angibt. Dies kann zu einer einheitlicheren Vergabe von Tags führen. Wenn „Auto“ 5000 Mal

vergeben wurde und „Autos“ nur 500 Mal, werden die meisten Nutzer höchwahrscheinlich eher „Auto“ als Tag wählen.

5.7 Flash Tutorials

Um die Hilfestellungen für Nutzer etwas zu erweitern, wird empfohlen, Flash Tutorials zu erstellen, ähnlich den Tutorials von Bibsonomy [7]. In diesen Tutorials könnte man u. a. die Funktion des Taggings erklären und zeigen, wie man Tags eingibt. Auch alle anderen Aspekte von Lycos iQ, vom Fragestellen bis zur Suche, könnten in Tutorials erklärt werden. Tutorials können Nutzer, die sich nicht so gerne solche Texte durchlesen, dazu bringen, sich dennoch den wichtigsten Inhalt anzusehen. Durch die visuelle Unterstützung solcher Flash Tutorials werden die Hilfestellungen interessanter, als wenn sie nur aus Text bestehen würden.

5.8 Rechtschreibhilfe

Es kommt vor, dass sich Nutzer bei der Eingabe von Tags vertippen, dies nicht bemerken und ihre Beiträge dadurch schlechter auffindbar werden. Dadurch wird auch die Gesamtqualität des Taggings verschlechtert. Dieses Problem könnte durch das Anfügen einer Rechtschreibhilfe gelöst werden. Wenn ein Wort eingegeben wird, welches von der Rechtschreibhilfe nicht erkannt wird, sollte ein Fenster erscheinen, das den Nutzer fragt, ob er wirklich dieses Wort als Tag vergeben will. So könnten Namen oder Abkürzungen trotzdem als Tags verwendet werden.

5.9 Alternative zur Rechtschreibhilfe

Wenn die Kosten oder der technische Aufwand für ein solches Wörterbuch mit Rechtschreibprüfung zu hoch sind, wäre es eine Alternative, die Lycos iQ Community einzubinden, indem man einigen Nutzern einen Moderatorenstatus verleiht, der sie u. a. dazu ermächtigt, Tags anderer zu verändern. Die Nutzer, die als Moderatoren fungieren, müssen natürlich vertrauenswürdig sein, damit sie ihre neuen Rechte nicht missbrauchen. So müssten die Moderatoren bestimmte Regeln beachten, z.B. dürften Tags nur gelöscht oder geändert werden, wenn sie eindeutig unsinnig bzw. falsch sind.

So kann es vorkommen, dass ein Nutzer bei seiner Frage als Tag „abcd“ eingibt, nur um ein Tag eingegeben zu haben, aber dieses Tag offensichtlich sinnlos ist. Wenn der Moderator allerdings selber nicht sicher ist, ob das Tag vielleicht ein Fachwort oder Akronym ist, darf er es nicht löschen oder ändern. Um Nutzer, deren Tags verändert werden, nicht zu irritieren, wäre es eine Möglichkeit, ihnen die Tags nur zu Änderung vorzuschlagen, damit diese die Änderungen selber per Klick bestätigen. Eine weitere Möglichkeit wäre, einem Moderator ein Themengebiet zu geben, welches er verwaltet. In Verbindung zu dem Kategorien-Vorschlag (vgl. Kap. 5.10), könnte ein Moderator so eine oder mehrere Kategorien zugeteilt bekommen, die zu seinem Interessengebiet passen.

Mit dieser Lösung würde man zwar nicht alle fehlerhaften oder sinnlosen Tags berichtigen, aber zumindest einen Teil davon. Die Qualität der Verschlagwortung mag nur geringfügig verbessert werden, allerdings würde diese Lösung ohne zusätzliche Kosten für Lycos iQ zustande kommen und könnte eine zusätzliche

Motivation für einige Nutzer bedeuten, die Moderatoren werden wollen. Sie würde sich auch mit dem Community-Charakter von Lycos iQ gut ergänzen lassen.

5.10 Kategorien

Dieser Vorschlag würde eine starke Änderung der Seite nach sich ziehen. Um die Qualität der Verschlagwortung weiter zu verbessern, wäre eine Einbindung von Kategorien zusätzlich zu den Tags sinnvoll, ähnlich wie es bei Fragr der Fall ist. Die Nutzer müssten also zu ihren Tags auch noch eine oder mehrere Kategorien für ihre Frage wählen. Es würde eine weitere Suchmöglichkeit angelegt werden müssen, eine Linkliste der Kategorien, über die man auf die einzelnen Kategorien zugreifen kann. Außerdem wäre es sinnvoll, dass neue wie alte Nutzer ihre Themengebiete, die man bei der Registrierung angeben muss, zukünftig nicht mehr frei definieren dürften, sondern aus den Kategorien auswählen müssten.

Die Kategorien dürften nicht zu allgemein sein, aber dennoch allgemein genug, um ein Themengebiet zu umfassen. Kategorien wie „alles und nichts“, welche bei Fragr benutzt wird, wären also nicht passend, da in eine solche Kategorie fast jede Frage passen würde. Eine Liste von elf Kategorien, welche aus dem Beitrag „Untersuchung der Qualität der Antworten bei Lycos iQ und deren Einbindung in die algorithmische Suche“ in diesem Band übernommen wurden, soll als ein Beispiel für mögliche Kategorien dienen.

- Musik
- Computer & Technisches
- Wissenschaften
- Medizin/Gesundheit
- Freizeit
- Filme
- Sprachen/Länder
- Tiere & Natur
- Beziehung / Liebe / Zwischenmenschliches
- Gesellschaft
- Sonstiges

Die Kategorie „Sonstiges“ ist für alle Fragen gedacht, die in keine der anderen Kategorien passen. Zusätzlich sollte man den Nutzern ein Mittel geben, um selbst Kategorien vorzuschlagen, die es noch nicht gibt. Dies könnte per Link auf der Eingabemaske geschehen, unter dem Feld, in dem dann die Kategorien ausgewählt werden können. Ein Beispiel wäre: „Keine passende Kategorie? Schlagen Sie eine vor!“ mit einem Link zu einem Formular, in welches dann der Vorschlag abgegeben werden kann. Wenn dann genügend Nutzer ähnliche Vorschläge gemacht haben, könnte man die neue Kategorie hinzufügen. Kategorien, die nicht oder sehr wenig benutzt werden, könnte man auch löschen, evtl. auch per Votum der Nutzer.

Das Einfügen von Kategorien zusätzlich zu den Tags würde das System einheitlicher machen, dennoch die Individualität des freien Taggings beibehalten. Nutzer könnten schneller ein bestimmtes Themengebiet finden bzw. Fragen in einem solchen Gebiet. Dennoch können Nutzer Tags vergeben, die ihre Frage nochmal

speziell auszeichnen. Durch die Kategorien würden wahrscheinlich weniger Nutzer allzu allgemeine Tags benutzen, z.B. Sport, Politik usw., da es diese ja bereits als Kategorien gibt. Es würden also etwas speziellere Tags verwendet werden, was die Qualität und Auffindbarkeit der Tags verbessern könnte.

6. FAZIT

Lycos iQ bietet bereits viele gängige Lösungen an, wie beispielsweise das Anbieten von Vorschlägen durch das „suggestive tagging“ oder das Anzeigen von verwandten Begriffen bei einer Suche. Nach der Analyse des Taggings anderer Dienste wurden einige Vorschläge entwickelt, die entweder bei anderen Anbietern so umgesetzt werden oder zumindest zu Ideen inspiriert haben.

Einige Anregungen sind einfacher umsetzbar, wie z.B. die Einbindung eines Videotutorials, um den Usern das Tagging zu erklären. Andere Vorschläge sind vermutlich aufwendiger, wie die Einbindung von Kategorien, die den Charakter der Seite stark ändern würde, aber dafür sorgen, dass keine zu allgemeinen Tags vergeben werden und eine Suche über Kategorien möglich wäre. Eine Erweiterung der Tagging-Rechte, so dass diejenigen, die eine Antwort geben, ebenfalls die Frage bzw. ihre Antwort taggen dürfen, würde helfen, dass Fragen umfangreicher und besser getaggt werden. Zudem wäre eine Kontrollinstanz eingebaut, die Tags auf Schreibfehler prüft und dem Fragenden zudem ein Feedback gibt. Zusätzliche Bearbeitungsmöglichkeiten, wie das Anhängen von Definitionen an eigenwillige oder individuelle Tags einzelner User, oder eine Erweiterung für die Editierung der Tags bieten dem User zusätzliche Features und damit eine neue Qualität und Ausweitung der Anwendungsmöglichkeiten.

Wenn die Fragen von den Nutzern besser verschlagwortet werden, dann sind sie folglich auch besser zu finden. Dies kann bei der Implementierung von Fragen und Antworten in klassischen Suchmaschinen helfen, denn wenn die Qualität des Taggings verbessert wird, könnten auch die in der Suchmaschinensuche angezeigten Fragen eine größere Relevanz hinsichtlich des eingegebenen Stichwortes haben.

Verwendete Literatur

1. del.icio.us (o.J.): What are Tags? Onlinedokument: <http://del.icio.us/help/tags> [Abruf am 19.12.2007]
2. Richter, Alexandra/Koch, Michael (2007): Social Software – Status quo und Zukunft Fakultät für Informatik, Universität der Bundeswehr München, Bericht Nr. 2007-01. Onlinedokument: http://www.unibw.de/wow5_3/forschung/social_software [Abruf am 19.12.2007]
3. Peters, Isabella (2007): Folksonomies im Web 2.0. In: Password, H. 5, S. 10-11.
4. Marlow, Cameron/Naaman, Mor/Boyd, Danah/Davis, Marc (2006): Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. Collaborative Web Tagging Workshop at WWW2006, Onlinedokument: <http://www.danah.org/papers/WWW2006.pdf> [Abruf am 19.12.2007]
5. Nguyen, Nick (o.J.): Profil von osunick bei del.icio.us. Onlinedokument: <http://del.icio.us/osunick> [Abruf am 19.12.2007]
6. Connotea (o.J.): Entity Describer. Onlinedokument: <http://www.connotea.org/wiki/EntityDescriber> [Abruf am 19.12.2007]
7. BibSonomy: Flash Tutorials. Onlinedokument: <http://www.bibsonomy.org/help/tutorials> [Abruf am 19.12.2007]

